

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2004-533829

(P2004-533829A)

(43) 公表日 平成16年11月11日(2004.11.11)

(51) Int.Cl. ⁷		F I		テーマコード (参考)	
C 1 2 N	15/09	C 1 2 N	15/00	Z N A A	2 G O 4 5
A 6 1 K	38/00	A 6 1 K	39/395	D	4 B O 2 4
A 6 1 K	39/395	A 6 1 K	39/395	N	4 B O 2 9
A 6 1 K	45/00	A 6 1 K	45/00		4 B O 5 0
A 6 1 P	1/04	A 6 1 P	1/04		4 B O 6 3
		審査請求 未請求 予備審査請求 未請求		(全 396 頁) 最終頁に続く	
(21) 出願番号	特願2003-500201 (P2003-500201)	(71) 出願人	301005050	インサイト・ゲノミックス・インコーポレイテッド アメリカ合衆国カリフォルニア州94304・パロアルト・ポータードライブ 3160 60	
(86) (22) 出願日	平成14年4月5日 (2002.4.5)	(74) 代理人	100089266		
(85) 翻訳文提出日	平成15年10月2日 (2003.10.2)		弁理士 大島 陽一		
(86) 国際出願番号	PCT/US2002/011152	(72) 発明者	アジムザイ、ヤルダ		
(87) 国際公開番号	W02002/097032		アメリカ合衆国カリフォルニア州94552・カストロバレー・ボールダーキャニオンドライブ 5518	最終頁に続く	
(87) 国際公開日	平成14年12月5日 (2002.12.5)				
(31) 優先権主張番号	60/282, 110				
(32) 優先日	平成13年4月6日 (2001.4.6)				
(33) 優先権主張国	米国 (US)				
(31) 優先権主張番号	60/283, 294				
(32) 優先日	平成13年4月11日 (2001.4.11)				
(33) 優先権主張国	米国 (US)				
(31) 優先権主張番号	60/286, 820				
(32) 優先日	平成13年4月26日 (2001.4.26)				
(33) 優先権主張国	米国 (US)				

(54) 【発明の名称】細胞成長、分化および細胞死関連タンパク質

(57) 【要約】

本発明は、ヒト細胞成長、分化および細胞死関連タンパク質(CGDD)および、CGDDを同定しコードするポリヌクレオチドを提供する。本発明はまた、発現ベクター、宿主細胞、抗体、アゴニストおよびアンタゴニストをも提供する。本発明はまた、CGDDの異常発現に関連する疾患を診断、治療または予防する方法をも提供する。

【特許請求の範囲】

【請求項 1】

以下の (a) 乃至 (i) からなる群から選択した単離されたポリペプチド。

(a) SEQ ID NO:1-21 (配列番号 1 乃至 21) からなる群から選択した或るアミノ酸配列を持つポリペプチド

(b) SEQ ID NO:3-4, SEQ ID NO:6, SEQ ID NO:8-9, SEQ ID NO:11-12, SEQ ID NO:14-16, SEQ ID NO:18-21 からなる群から選択した或るアミノ酸配列に対して少なくとも 90% が同一であるような天然アミノ酸配列を有するポリペプチド

(c) SEQ ID NO:1 のアミノ酸配列に対して少なくとも 93% が同一であるような天然アミノ酸配列を有するポリペプチド

(d) SEQ ID NO:2 のアミノ酸配列に対して少なくとも 95% が同一であるような天然アミノ酸配列を有するポリペプチド

(e) SEQ ID NO:5 のアミノ酸配列に対して少なくとも 96% が同一であるような天然アミノ酸配列を有するポリペプチド

(f) SEQ ID NO:7 および SEQ ID NO:10 からなる群から選択した或るアミノ酸配列に対して少なくとも 98% が同一であるような天然アミノ酸配列を有するポリペプチド

(g) SEQ ID NO:17 のポリヌクレオチド配列に対して少なくとも 99% が同一であるような天然アミノ酸配列を有するポリペプチド

(h) SEQ ID NO:1-21 からなる群から選択したアミノ酸配列を有するポリペプチドの生物学的活性断片

(i) SEQ ID NO:1-21 からなる群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片

【請求項 2】

SEQ ID NO:1-21 からなる群から選択したアミノ酸配列を有する、請求項 1 に記載の単離されたポリペプチド。

【請求項 3】

請求項 1 のポリペプチドをコードする単離されたポリヌクレオチド。

【請求項 4】

請求項 2 のポリペプチドをコードする単離されたポリヌクレオチド。

【請求項 5】

SEQ ID NO:22-42 からなる群から選択したポリヌクレオチド配列を有する、請求項 4 に記載の単離されたポリヌクレオチド。

【請求項 6】

請求項 3 に記載のポリヌクレオチドへ機能的に連結したプロモーター配列を有する組換えポリヌクレオチド。

【請求項 7】

請求項 6 に記載の組換えポリヌクレオチドを用いて形質転換した細胞。

【請求項 8】

請求項 6 に記載の組換えポリヌクレオチドを有する遺伝形質転換体。

【請求項 9】

請求項 1 のポリペプチドを生産する方法であって、

(a) 前記ポリペプチドの発現に好適な条件下で、請求項 1 のポリペプチドをコードする或るポリヌクレオチドへ機能的に連結されたプロモーター配列を持つ組換えポリヌクレオチドで形質転換される細胞を培養する過程と、

(b) そのように発現した前記ポリペプチドを回収する過程、とからなることを特徴とする方法。

【請求項 10】

前記ポリペプチドが、SEQ ID NO:1-21 からなる群から選択した或るアミノ酸配列を含むことを特徴とする、請求項 9 に記載の方法。

【請求項 11】

10

20

30

40

50

請求項 1 に記載のポリペプチドと特異結合するような単離された抗体。

【請求項 1 2】

以下の (a) 乃至 (k) からなる群から選択した単離されたポリヌクレオチド。

- (a) SEQ ID NO:22-42 からなる群から選択した或るポリヌクレオチド配列を有するポリヌクレオチド
- (b) SEQ ID NO:22-25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:32-37, SEQ ID NO:39-42 からなる群から選択した或るアミノ酸配列に対して少なくとも 90% が同一であるような天然アミノ酸配列を有するポリペプチド
- (c) SEQ ID NO:26 のポリヌクレオチド配列に対して少なくとも 92% が同一であるような天然ポリヌクレオチド配列を有するポリヌクレオチド
- (d) SEQ ID NO:28, SEQ ID NO:31 からなる群から選択した或るポリヌクレオチド配列に対して少なくとも 98% が同一であるような天然ポリヌクレオチド配列を有するポリヌクレオチド
- (e) SEQ ID NO:38 のポリヌクレオチド配列に対して少なくとも 99% が同一であるような天然ポリヌクレオチド配列を有するポリヌクレオチド
- (f) (a) のポリヌクレオチドに相補的なポリヌクレオチド
- (g) (b) のポリヌクレオチドに相補的なポリヌクレオチド
- (h) (c) のポリヌクレオチドに相補的なポリヌクレオチド
- (i) (d) のポリヌクレオチドに相補的なポリヌクレオチド
- (j) (e) のポリヌクレオチドに相補的なポリヌクレオチド
- (k) (a) ~ (j) の RNA 等価物

10

20

【請求項 1 3】

請求項 12 に記載のポリヌクレオチドの少なくとも 60 の連続したヌクレオチドを持つ、単離されたポリヌクレオチド。

【請求項 1 4】

請求項 12 に記載のポリヌクレオチドの配列を有する標的ポリヌクレオチドをサンプル中から検出する方法であって、

- (a) 前記サンプル中の前記標的ポリヌクレオチドに相補的な或る配列を有する少なくとも 20 の連続したヌクレオチド群を持つ或るプローブと前記サンプルとをハイブリダイズし、該プローブが、或るハイブリダイゼーション複合体が前記プローブと前記標的ポリヌクレオチドまたはその断片群との間に形成される条件下で、前記標的ポリヌクレオチドへ特異的にハイブリダイズする過程と、
- (b) 前記ハイブリダイゼーション複合体の有無を検出し、該複合体が存在する場合にはオプションでその量を検出する過程、とを含む方法。

30

【請求項 1 5】

前記プローブが少なくとも 60 の連続したヌクレオチドを含むことを特徴とする請求項 14 に記載の方法。

【請求項 1 6】

請求項 12 に記載のポリヌクレオチドの配列を有する標的ポリヌクレオチドをサンプル中から検出する方法であって、

40

- (a) ポリメラーゼ連鎖反応増幅を用いて前記標的ポリヌクレオチドまたはその断片を増幅する過程と、
- (b) 前記増幅した標的ポリヌクレオチドまたはその断片の有無を検出し、該標的ポリヌクレオチドまたはその断片が存在する場合にはオプションでその量を検出する過程、とを含むことを特徴とする方法。

【請求項 1 7】

請求項 1 のポリペプチドと、薬剤として許容できる賦形剤とを含むことを特徴とする組成物。

【請求項 1 8】

前記ポリペプチドが、SEQ ID NO:1-21 からなる群から選択したアミノ酸配列を含むことを

50

特徴とする請求項17の組成物。

【請求項19】

機能的CGDDの発現低下に関連する疾患又は病状を治療する方法であって、そのような治療を必要とする患者に対して請求項17に記載の組成物を投与する過程を含むことを特徴とする方法。

【請求項20】

請求項1のポリペプチドのアゴニストとしての有効性を確認するために化合物をスクリーニングする方法であって、

(a) 請求項1のポリペプチドを有する或るサンプルを或る化合物に曝す過程と、

(b) 前記サンプルにおいてアゴニスト活性を検出する過程、とを含むことを特徴とするスクリーニング方法。 10

【請求項21】

請求項20に記載の方法によって同定したアゴニスト化合物と、薬剤として許容できる賦形剤とを含むことを特徴とする組成物。

【請求項22】

機能的CGDDの発現低下に関連する疾患又は病状を治療する方法であって、そのような治療を必要とする患者に対して請求項21に記載の組成物を投与する過程を含むことを特徴とする方法。

【請求項23】

請求項1のポリペプチドのアンタゴニストとしての有効性を確認するために化合物をスクリーニングする方法であって、 20

(a) 請求項1のポリペプチドを有する或るサンプルを或る化合物に曝す過程と、

(b) 前記サンプルにおいてアンタゴニスト活性を検出する過程、とを含むことを特徴とするスクリーニング方法。

【請求項24】

請求項23の方法で同定したアンタゴニスト化合物と、薬剤として許容できる賦形剤とを有する組成物。

【請求項25】

機能的CGDDの過剰な発現に関連する疾患や病態の治療方法であって、そのような治療を要する患者への請求項24の組成物の投与を含む治療方法。 30

【請求項26】

請求項1のポリペプチドに特異結合する化合物をスクリーニングする方法であって、

(a) 請求項1に記載のポリペプチドを適切な諸条件下で少なくとも1つの試験化合物と混合する過程と、

(b) 請求項1に記載のポリペプチドの試験化合物との結合を検出し、それによって請求項1に記載のポリペプチドに特異結合する化合物を同定する過程、とを含むことを特徴とする方法。

【請求項27】

請求項1のポリペプチドの活性をモジュレート（調節）する化合物をスクリーニングする方法であって、 40

(a) 請求項1に記載のポリペプチドの活性が許容された諸条件下で、請求項1に記載のポリペプチドを少なくとも1つの試験化合物と混合する過程と、

(b) 請求項1に記載のポリペプチドの活性を試験化合物の存在下で算定する過程と、

(c) 試験化合物の存在下での請求項1に記載のポリペプチドの活性を、試験化合物の不存在下での請求項1に記載のポリペプチドの活性と比較する過程とを含み、試験化合物の存在下での請求項1に記載のポリペプチドの活性の変化が、請求項1に記載のポリペプチドの活性を調節する化合物を標示することを特徴とする方法。

【請求項28】

請求項5の配列を持つ標的ポリヌクレオチドの発現を改変するのに効果的な化合物をスクリーニングする方法であって、 50

- (a) 前記標的ポリヌクレオチドの発現に好適な諸条件下で、前記標的ポリヌクレオチドを有する或るサンプルを或る化合物に曝露する過程と、
- (b) 前記標的ポリヌクレオチドの、改変された発現を検出する過程と、
- (c) 可変量の前記化合物の存在下と前記化合物の不存在下で、前記標的ポリヌクレオチドの発現を比較する過程、とを含むことを特徴とする方法。

【請求項 29】

試験化合物の毒性を算定する方法であって、

- (a) 核酸群を有する或る生体サンプルを前記試験化合物で処理する過程と、
- (b) 処理した前記生体サンプルの核酸群と、請求項12のポリヌクレオチドの少なくとも20の連続するヌクレオチド群を有する或るプローブをハイブリダイズさせる過程であって、このハイブリダイゼーションが、前記プローブと前記生体サンプルの或る標的ポリヌクレオチドとの間で或る特異的なハイブリダイゼーション複合体が形成される諸条件下で行われ、前記標的ポリヌクレオチドが、請求項12のポリヌクレオチドのポリヌクレオチド配列またはその断片を有するポリヌクレオチドである、前記過程と、
- (c) ハイブリダイゼーション複合体の収量を定量する過程と、
- (d) 前記処理された生体サンプル中の前記ハイブリタイゼーション複合体の量を、或る処理されていない生体サンプル中の前記ハイブリタイゼーション複合体の量と比較する過程とを含み、前記処理された生体サンプル中の前記ハイブリタイゼーション複合体の量の差が、前記試験化合物の毒性を標示することを特徴とする方法。

【請求項 30】

生物学的サンプル中のCGDDの発現に関連する症状または疾患に対する診断試験法であって、

- (a) 前記抗体が前記ポリペプチドに結合し、抗体とポリペプチドとの複合体が形成されるのに適した諸条件下で、前記生体サンプルを請求項11に記載の抗体と混合する過程と、
- (b) 前記複合体を検出する過程とを含み、前記複合体の存在が、前記生体サンプル中の前記ポリペプチドの存在と相関することを特徴とする方法。

【請求項 31】

請求項11の抗体であって、

- (a) キメラ抗体
- (b) 一本鎖抗体
- (c) Fab断片
- (d) F(ab')₂ 断片
- (e) ヒト化抗体、のいずれかであることを特徴とする抗体。

【請求項 32】

請求項11に記載の抗体と、許容できる賦形剤とを有する組成物。

【請求項 33】

CGDDの被検者での発現に関連する症状又は疾患の診断方法であって、請求項32に記載の組成物の有効量を前記被検者に投与する過程を含むことを特徴とする方法。

【請求項 34】

前記抗体が標識されることを特徴とする、請求項32に記載の組成物。

【請求項 35】

CGDDの被検者での発現に関連する症状又は疾患の診断方法であって、請求項34に記載の組成物の有効量を前記被検者に投与する過程を含むことを特徴とする方法。

【請求項 36】

請求項11の抗体の特異性を持つポリクローナル抗体を調製する方法であって、

- (a) 抗体応答を誘発する諸条件下で、SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列またはその免疫原性断片を有する或るポリペプチドを用いて或る動物を免疫化する過程と、
- (b) 前記動物から抗体を単離する過程と、
- (c) 前記単離された抗体を該ポリペプチドでスクリーニングし、それによって、SEQ ID

N0:1-21からなる群から選択した或るアミノ酸配列を有する或るポリペプチドに特異結合するようなポリクローナル抗体を同定する過程、とを含むことを特徴とする方法。

【請求項 37】

請求項36の方法で産生したポリクローナル抗体。

【請求項 38】

請求項37のポリクローナル抗体と好適なキャリアとを有する組成物。

【請求項 39】

請求項11に記載の抗体の特異性を有するモノクローナル抗体を作製する方法であって、

(a) 抗体応答を誘発する諸条件下で、SEQ ID N0:1-21からなる群から選択した或るアミノ酸配列またはその免疫原性断片を有する或るポリペプチドを用いて或る動物を免疫化する過程と、 10

(b) 前記動物から、抗体を産出する細胞を単離する過程と、

(c) 不死化した細胞と前記抗体産出細胞とを融合し、モノクローナル抗体を産出するハイブリドーマ細胞を形成する過程と、

(d) 前記ハイブリドーマ細胞を培養する過程と、

(e) SEQ ID N0:1-21からなる群から選択したアミノ酸配列を有するポリペプチドに特異結合するようなモノクローナル抗体を前記培養物から単離する過程、とを含むことを特徴とする方法。

【請求項 40】

請求項39に記載の方法で産出したモノクローナル抗体。 20

【請求項 41】

請求項40に記載のモノクローナル抗体と適切なキャリアとを有する組成物。

【請求項 42】

Fab発現ライブラリのスクリーニングによって前記抗体を産出することを特徴とする、請求項11に記載の抗体。

【請求項 43】

組換え免疫グロブリンライブラリをスクリーニングすることにより産出されることを特徴とする、請求項11に記載の抗体。

【請求項 44】

SEQ ID N0:1-21からなる群から選択したアミノ酸配列を有するポリペプチドをサンプル中に検出する方法であって、 30

(a) 前記抗体と前記ポリペプチドとの特異結合を許容する条件下で、或るサンプルと共に請求項11に記載の抗体をインキュベートする過程と、

(b) 特異結合を検出する過程とを含み、該特異結合が、SEQ ID N0:1-21からなる群から選択したアミノ酸配列を有するポリペプチドがサンプル中に存在することを標示することを特徴とする方法。

【請求項 45】

SEQ ID N0:1-21 からなる群から選択したアミノ酸配列を有するポリペプチドを精製する方法であって、

(a) 前記抗体と前記ポリペプチドとの特異結合を許容する条件下で、或るサンプルと共に請求項11に記載の抗体をインキュベートする過程と、 40

(b) 前記サンプルから前記抗体を分離し、SEQ ID N0:1-21からなる群から選択した或るアミノ酸配列を有する精製ポリペプチドを得る過程、とを含むことを特徴とする方法。

【請求項 46】

マイクロアレイの少なくとも1つのエレメントが請求項13に記載のポリヌクレオチドであることを特徴とするマイクロアレイ。

【請求項 47】

ポリヌクレオチド群を有する或るサンプルの発現プロフィールを作製する方法であって、

(a) 該サンプルの該ポリヌクレオチド群を標識する過程と、

(b) ハイブリダイゼーション複合体の形成に適した諸条件下で請求項46のマイクロアレ 50

イのエレメント群をサンプルの標識されたポリヌクレオチド群と接触させる過程と、
(c) 該サンプル中の該ポリヌクレオチド群の発現を定量化する過程、とを含むことを特徴とする方法。

【請求項 4 8】

或る固体基板上の固有の物理的位置に付着された種々のヌクレオチド分子を有するアレイであって、少なくとも 1 つの前記ヌクレオチド分子が、或る標的ポリヌクレオチドの少なくとも 30 の連続したヌクレオチド群と特異的にハイブリダイズ可能な最初のオリゴヌクレオチドまたはポリヌクレオチド配列を含み、前記の標的ポリヌクレオチドが請求項 12 に記載のポリヌクレオチドであることを特徴とするアレイ。

【請求項 4 9】

請求項 48 に記載のアレイで、前記の最初のオリゴヌクレオチドまたはポリヌクレオチドの配列が前記の標的ポリヌクレオチドの少なくとも 30 の連続したヌクレオチドに完全に相補的であることを特徴とするアレイ。

【請求項 5 0】

請求項 48 に記載のアレイで、前記の最初のオリゴヌクレオチドまたはポリヌクレオチドの配列が前記の標的ポリヌクレオチドの少なくとも 60 の連続したヌクレオチドに完全に相補的であることを特徴とするアレイ。

【請求項 5 1】

請求項 48 に記載のアレイで、前記のオリゴヌクレオチドまたはポリヌクレオチドの最初の配列が前記の標的ポリヌクレオチドに完全に相補的であることを特徴とするアレイ。

【請求項 5 2】

請求項 48 に記載のアレイで、マイクロアレイであることを特徴とするアレイ。

【請求項 5 3】

請求項 48 に記載のアレイで、前記のオリゴヌクレオチドまたはポリヌクレオチドの最初の配列を有するヌクレオチド分子にハイブリダイズした前記の標的ポリヌクレオチドを有することを特徴とするアレイ。

【請求項 5 4】

請求項 48 に記載のアレイで、或るリンカーが少なくとも 1 つの前記のヌクレオチド分子と前記の固体基板とを連結していることを特徴とするアレイ。

【請求項 5 5】

請求項 48 に記載のアレイで、該基板上の固有の物理的位置の各々が複数のヌクレオチド分子を含み、任意の単一の固有の物理的位置でのその複数のヌクレオチド分子は同一の配列を有し、該基板上の固有の物理的位置の各々は、該基板上の別の固有の物理的位置でのヌクレオチド分子群の配列とは異なる或る配列を有するヌクレオチド分子群を含むことを特徴とするアレイ。

【請求項 5 6】

SEQ ID N0:1 のアミノ酸配列を有する請求項 1 に記載のポリペプチド。

【請求項 5 7】

SEQ ID N0:2 のアミノ酸配列を有する請求項 1 に記載のポリペプチド。

【請求項 5 8】

SEQ ID N0:3 のアミノ酸配列を有する請求項 1 に記載のポリペプチド。

【請求項 5 9】

SEQ ID N0:4 のアミノ酸配列を有する請求項 1 に記載のポリペプチド。

【請求項 6 0】

SEQ ID N0:5 のアミノ酸配列を有する請求項 1 に記載のポリペプチド。

【請求項 6 1】

SEQ ID N0:6 のアミノ酸配列を有する請求項 1 に記載のポリペプチド。

【請求項 6 2】

SEQ ID N0:7 のアミノ酸配列を有する請求項 1 に記載のポリペプチド。

【請求項 6 3】

10

20

30

40

50

SEQ ID NO:8のアミノ酸配列を有する請求項1に記載のポリペプチド。

【請求項64】

SEQ ID NO:9のアミノ酸配列を有する請求項1に記載のポリペプチド。

【請求項65】

SEQ ID NO:10のアミノ酸配列を有する請求項1に記載のポリペプチド。

【請求項66】

SEQ ID NO:11のアミノ酸配列を有する請求項1に記載のポリペプチド。

【請求項67】

SEQ ID NO:12のアミノ酸配列を有する請求項1に記載のポリペプチド。

【請求項68】

SEQ ID NO:13のアミノ酸配列を有する請求項1に記載のポリペプチド。

【請求項69】

SEQ ID NO:14のアミノ酸配列を有する請求項1に記載のポリペプチド。

【請求項70】

SEQ ID NO:15のアミノ酸配列を有する請求項1に記載のポリペプチド。

【請求項71】

SEQ ID NO:16のアミノ酸配列を有する請求項1に記載のポリペプチド。

【請求項72】

SEQ ID NO:17のアミノ酸配列を有する請求項1に記載のポリペプチド。

【請求項73】

SEQ ID NO:18のアミノ酸配列を有する請求項1に記載のポリペプチド。

【請求項74】

SEQ ID NO:19のアミノ酸配列を有する請求項1に記載のポリペプチド。

【請求項75】

SEQ ID NO:20のアミノ酸配列を有する請求項1に記載のポリペプチド。

【請求項76】

SEQ ID NO:21のアミノ酸配列を有する請求項1に記載のポリペプチド。

【請求項77】

SEQ ID NO:22のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項78】

SEQ ID NO:23のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項79】

SEQ ID NO:24のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項80】

SEQ ID NO:25のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項81】

SEQ ID NO:26のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項82】

SEQ ID NO:27のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項83】

SEQ ID NO:28のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項84】

SEQ ID NO:29のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項85】

SEQ ID NO:30のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項86】

SEQ ID NO:31のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項87】

SEQ ID NO:32のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項88】

10

20

30

40

50

SEQ ID NO:33のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項 8 9】

SEQ ID NO:34のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項 9 0】

SEQ ID NO:35のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項 9 1】

SEQ ID NO:36のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項 9 2】

SEQ ID NO:37のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項 9 3】

SEQ ID NO:38のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項 9 4】

SEQ ID NO:39のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項 9 5】

SEQ ID NO:40のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項 9 6】

SEQ ID NO:41のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【請求項 9 7】

SEQ ID NO:42のポリヌクレオチド配列を有する請求項12に記載のポリヌクレオチド。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、細胞成長、分化および細胞死関連タンパク質の核酸配列およびアミノ酸配列に関する。本発明はまた、これらの配列を利用した、癌を含む細胞増殖異常、発生または発達障害、神経疾患、自己免疫または炎症性疾患、生殖障害、および胎盤の障害の、診断・治療・予防に関する。本発明はさらに、細胞成長、分化および細胞死関連タンパク質の核酸配列およびアミノ酸配列の発現における外因性化合物の効果についての評価に関する。

【背景技術】

【0002】

ヒトの成長と発達には、細胞分化、細胞増殖、およびアポトーシスの、空間的および時間的な調節を要する。これらのプロセスが、生殖、加齢、胚形成、形態形成、器官形成、並びに組織の修復および維持を、協調的に制御する。細胞レベルでは、成長と発達との支配は、細胞が細胞分裂周期から出たり入ったりする決定に、また、細胞の、最終分化状態へのコミットメントによって成される。これらの決定は、細胞によって、細胞外シグナルと他の細胞が受け取る環境からの合図に応じて成される。以下の考察の焦点は、細胞分裂、胚形成、細胞分化および細胞増殖、並びにアポトーシスの分子機序に置かれ、また、これら機序の破壊に起因しうる癌などの病態に置かれている。

【0003】

細胞周期

細胞分裂は、全ての生物がこれによって成長し生殖する、基本的プロセスである。酵母菌や細菌などの単細胞生物では、各細胞分裂で生物の数が2倍となる。多細胞の種では、消耗あるいはプログラム細胞死により消失した細胞を置換するために、また新たな組織や器官を作製する細胞分化のために、多数回の細胞分裂が必要とされる。細胞周期の進行は、タンパク質複合体群の複雑な相互作用によって支配される。この調節は、成長因子や他の有糸分裂促進因子(マイトジェン)などの細胞外信号と、DNA損傷または栄養飢餓などの細胞内合図とに応じて細胞周期の進行を制御するタンパク質の、適切な発現に依存する。細胞周期の進行を直接／間接的に調節する分子は、サイクリン、サイクリン依存性プロテインキナーゼ、成長因子とその受容体、セカンドメッセンジャー、シグナル伝達タンパク質、発癌遺伝子産物、腫瘍抑制タンパク質等の複数のカテゴリーに分類される。

【0004】

10

20

30

40

50

細胞周期の進行は、タンパク質複合体群の複雑な相互作用によって支配される。この調節は、成長因子や他の有糸分裂促進因子(マイトジェン)などの細胞外信号と、DNA損傷または栄養飢餓などの細胞内合図とに応じて細胞周期の進行を制御するタンパク質の、適切な発現に依存する。細胞周期の進行を直接／間接的に調節する分子は、サイクリン、サイクリン依存性プロテインキナーゼ、成長因子とその受容体、セカンドメッセンジャー、シグナル伝達タンパク質、発癌遺伝子産物、腫瘍抑制タンパク質等の複数のカテゴリーに分類される。

【0005】

細胞は、サイクリンと呼ばれる活性化タンパク質ファミリーの合成および分解による調節により、有糸分裂に入ったり有糸分裂から出たりする。サイクリン是一群のサイクリン依存性プロテインキナーゼ(Cdk)へ結合し活性化させて作用する。Cdkは次に、有糸分裂プロセスに関与する選択されたタンパク質をリン酸化し活性化する。サイクリンは、長さが約180アミノ酸であり「サイクリンボックス」と呼ばれる、共有した相同性を持つ広い領域により特徴付けられる(Chapman, D.L及びWolgemuth, D.J. (1993) Development 118:229-40)。さらにサイクリン内には、「destruction box」と呼ばれる、この分子のN末端領域内に保存された9アミノ酸配列を含む。この配列は、ユビキチンが仲介するサイクリンB分解を誘発する認識コードと思われる(Hunt, T. (1991) Nature 349:100-101)。数タイプのサイクリンが存在する(Ciechanover, A. (1994) Cell 79:13-21)。G1期とS期との過程は、G1サイクリン類とその触媒サブユニット群(例えばCdk2-サイクリンA、Cdk2-サイクリンE、Cdk4-サイクリンDおよびCdk6-サイクリンDを含む)によって進行される。G2/Mの移行は、Cdc2サイクリンA、Cdc2サイクリンB1、Cdc2サイクリンB2の複合体等の、有糸分裂CDKサイクリン複合体の活性化によって進行される(Yang, J及びKornbluth, S. (1999) Trends in Cell Biology 9:207-210に概説)。

10

20

【0006】

サイクリンは、真核細胞と数種の細菌では、細胞タンパク質の分解の主要な経路である、ユビキチン抱合系(ubiquitin conjugation system:UCS)によって分解される。UCSは異常なタンパク質の除去を仲介する。また遺伝子の転写と細胞周期の進行等の細胞過程を制御する重要な調節タンパク質の半減期を調節する。UCSは、有糸分裂サイクリンキナーゼ、オンコプロテイン(腫瘍性タンパク)、腫瘍抑制遺伝子(p53等)、ウイルス蛋白、シグナル伝達に関係する細胞表面受容体、転写調節因子、および変異タンパク質または損傷タンパク質の分解に関与する(Ciechanover, 前出)。

30

細胞分裂周期の詳細は異なりうるが、基本的なプロセスは3つの主要な現象からなる。最初の現象である間期には、細胞分裂の準備、DNAの複製、必須タンパク質の作製が含まれる。2番目の現象である有糸分裂では、核物質が分裂して細胞の両側に分離する。最終の現象である細胞質分裂は、細胞質の分配と分裂である。細胞周期移行の順序およびタイミングは、種々のチェックポイント(検問点)で正または負の調節回路によりプロセスを制御する細胞周期調節システムの制御下にある。

【0007】

有糸分裂は間期の終了を特徴付け、また細胞質分裂の開始で終了する。有糸分裂には4つの段階があり、次の順番で発生する。前期、中期、後期、終期。前期には二極性紡錘体の形成が含まれ、これは微小管と、ダイニンなどの関連タンパク質から構成されており、形成は極性分裂中心から始まる。中期の間、核物質は凝縮する。また、核物質の、紡錘体への物理的な付着を助ける動原体線維を発達させる。続いて核物質は紡錘体に沿って反対の極へ移動するが、それは後期に起こる。終期には、核物質からの紡錘体と動原体線維の消滅が含まれる。有糸分裂は、多数のタンパク質の相互作用に依存する。例えば、CENP-A、B、C等のセントロメア結合タンパク質は、動原体の形成およびアセンブリにおいて構造的役割を果たしている(Saffery, R. 他(2000) Human Mol. Gen. 9: 175-185)。

40

【0008】

真核細胞周期のM期に構造的再編成が起こり、娘細胞間の細胞成分の適切な分配を確保する。間期構造の、より小さなサブユニットへの分解は一般的である。核エンベロープは小

50

胞へ分解し、核ラミンは解離する。次にこれらラミンのリン酸化が起こり、リン酸化は終期まで維持される。終期には、核ラミナ構造が再構成される。M期リン酸化(MPP)をコードするcDNA群はU3核小体低分子リボ核タンパク質(snoRNP)の成分であり、これらcDNAは、有糸分裂が完了すると、核小体に再局在する(Westendorf, J.M. 他(1998) J. Biol. Chem. 9:437-449)。U3 snoRNP類は、RNAプロセッシング諸イベントの必須メディエータである。

【0009】

細胞プロセス(例えば有糸分裂)の調節に関わるタンパク質としては、Ser/Thrプロテインホスファターゼ1型(PP-1)を含む。PP-1類の作用は、中期-後期移行に関わる、鍵となるタンパク質群の脱リン酸化によって成される。遺伝子PP1R7は調節ポリペプチドsds22をコードし、PP1R7は少なくとも6種のスプライス変異体を持つ(Ceulemans, H. 他(1999) Eur. J. Biochem. 262:36-42)。sds22はPP-1類の触媒サブユニットの活性を調節し、有糸分裂基質類の、PP-1依存性脱リン酸化を増進する。 10

【0010】

細胞周期調節タンパク質類は、細胞増殖と癌とに重要な役割を果たす。例えば、細胞周期イベントの妥当な実行とタイミングとが不全であると、染色体分配欠損をもたらすことがあり、この結果、異数性または倍数性が生じる。このゲノム不安定性は、癌化した細胞の特徴である(Luca, F.C. and Winey, M. (1998) Mol. Biol. Cell. 9:29-46)。最近同定されたタンパク質であるmMOB1は酵母MOB1の哺乳類相同体であり、MOB1は、有糸分裂の完遂と倍数関係の維持とに必要な必須の酵母遺伝子である。哺乳類mMOB1はプロテインホスファターゼ2A(PP2A)を含むタンパク質複合体類のメンバーであり、mMOB1のリン酸化は、PP2Aによって調節されるようである(Moreno, C.S. 他(2001) J. Biol. Chem. 276:24253-24260)。PP2Aは、ヒト癌の発症に関わるとされ、これには肺癌、結腸癌、および白血病を含む。 20

【0011】

細胞周期の調節には、連続して相互作用する、多数のタンパク質が関わる。真核生物の細胞周期は幾つかの、高度に制御されたイベントからなり、それらイベントの正確な秩序が、DNA複製と細胞分裂との成功を保証する。細胞は、これらイベントの秩序の維持を、後期イベントを早期イベントの成功裏の完了に依存させることによって行う。この依存性は、チェックポイントと呼ばれる細胞機序によって補強される。更なる細胞周期調節タンパク質の例には、ヒストンデアセチラーゼ(HDAC)類を含む。HDAC類は細胞周期の調節に関与する。また、クロマチン構造を調整する。ヒトHDAC1はin vitroでヒトHus1遺伝子産物と相互作用することが知られている。この産物の分裂酵母(Schizosaccharomyces pombe)相同体は、G₂/Mチェックポイント制御に関わると示唆されている(Cai, R.L. 他(2000) J. Biol. Chem. 275:27909-27916)。 30

【0012】

DNA損傷(G₂)とDNA複製(S期)とのチェックポイントは、真核細胞をG₂/M移行期に停止させる。この停止は、有糸分裂に入る前に、DNA修復またはDNA複製が起こるための時間を提供する。したがって、G₂/Mチェックポイントは、DNA複製が完了し、染色体損傷が無いときにのみ、有糸分裂が起こることを保証する。分裂酵母のHus1遺伝子は細胞周期チェックポイント遺伝子であり、radファミリーの遺伝子(例えばrad1およびrad9)も同様である(Volkmer, E. および Karnitz, L.M. (1999) J. Biol. Chem. 274:567-570; Kostrub C.F. 他(1998) EMBO J. 17:2055-2066)。これら遺伝子は有糸分裂チェックポイントに関わる。また、DNA損傷または複製の遮断のいずれかによってこれらの遺伝子が誘導される。DNA損傷または複製遮断の誘発は、Hus1遺伝子の機能の喪失と、続いての細胞死とをもたらす。大多数のrad遺伝子のヒト相同体群が同定されており、これにはATMとATRとを含み、これらはrad3pのヒト相同体である。ATM遺伝子での突然変異群は、重度の先天性疾患である血管拡張性失調症と相関する(Savitsky, K. 他(1995) Science 268:1749-1753)。ヒトHus1タンパク質はrad9と相互作用するrad1タンパク質との複合体の形で作用することが示されており、このことは、これらを、ヒト細胞の、或るDNA損傷応答性タンパク質複合体の中心成分にしている(Volkmer, E. および Karnitz, L.M. (1999) J. Biol. Chem. 274:567- 40 50

570。

【0013】

細胞は、サイクリンと呼ばれる活性化タンパク質ファミリーの合成および分解による調節により、有糸分裂に入ったり有糸分裂から出たりする。サイクリンは一群のサイクリン依存性プロテインキナーゼ(Cdk)へ結合し活性化させて作用する。Cdkは次に、有糸分裂プロセスに関与する選択されたタンパク質をリン酸化し活性化する。サイクリンは、長さが約180アミノ酸であり「サイクリンボックス」と呼ばれる、共有した相同性を持つ広い領域により特徴付けられる(Chapman, D.L及びWolgemuth, D.J. (1993) Development 118:229-40)。さらにサイクリン内には、「destruction box」と呼ばれる、この分子のN末端領域内に保存された9アミノ酸配列を含む。この配列は、ユビキチンが仲介するサイクリンB分解を誘発する認識コードと思われる(Hunt, T. (1991) Nature 349:100-101)。数タイプのサイクリンが存在する(Ciechanover, A. (1994) Cell 79:13-21)。G1期とS期との過程は、G1サイクリン類とその触媒サブユニット群(例えばCdk2-サイクリンA、Cdk2-サイクリンE、Cdk4-サイクリンDおよびCdk6-サイクリンDを含む)によって進行される。G2/Mの移行は、Cdc2サイクリンA、Cdc2サイクリンB1、Cdc2サイクリンB2複合体等の、有糸分裂CDKサイクリン複合体の活性化によって進行される(Yang, J及びKornbluth, S. (1999) Trends in Cell Biology 9:207-210に概説)。

10

【0014】

サイクリンは、真核細胞と数種の細菌では、細胞タンパク質の分解の主な経路である、ユビキチン抱合系(ubiquitin conjugation system:UCS)によって分解される。UCSは異常なタンパク質の除去を仲介する。また遺伝子の転写と細胞周期の進行等の細胞過程を制御する重要な調節タンパク質の半減期を調節する。UCSは、有糸分裂サイクリンキナーゼ、オンコプロテイン(腫瘍性タンパク)、腫瘍抑制遺伝子(p53等)、ウイルス蛋白、シグナル伝達に関係する細胞表面受容体、転写調節因子、および変異タンパク質または損傷タンパク質の分解に関与する(Ciechanover, 前出)。

20

【0015】

ユビキチン抱合とタンパク質分解のプロセスは5つの主要なステップの形で起こる(Jentsch, S. (1992) Annu. Rev. Genet. 26:179-207)。第1番目にATP依存性反応で、小さな耐熱性タンパク質であるユビキチン(Ub)がユビキチン活性化酵素(E1)により活性化する。この反応はUbのC末端をE1内部システイン残基のチオール基に結合する。2番目に、活性化したUbがUb抱合酵素(E2)のうちの1つに移る。異なるユビキチン依存性タンパク質分解経路は構造的に類似するが異なるユビキチン抱合酵素を利用し、これら酵素は特定の分解シグナルを運ぶタンパク質に酵素を誘導する認識サブユニット群と会合する。3番目にE2は、C末端グリシンによりユビキチン-タンパク質リガーゼファミリー、E3のメンバーにUb分子を移す。4番目に、E3はUb分子を標的タンパク質に移す。付加的Ub分子が標的タンパク質に追加され多重Ub鎖構造を形成することがある。5番目に、ユビキチン結合したタンパク質が次に、大きな多サブユニットタンパク質分解酵素複合体であるプロテアソームによって認識され分解される。そして再利用のためにUbが放出される。

30

【0016】

活性化の前に、Ubは通常、融合タンパク質(N末端のユビキチンと、C末端延長タンパク質(CEP)で構成)として発現、または、ポリユビキチン蛋白質(Ub単量体群が同じ方向を向いて一列(head to tail)に結合)として発現される。CEP類は種々の調節タンパク質の特徴を持つ。即ち大抵は高度に塩基性であり、最大30%のリジン残基とアルギニン残基とを含み、また、核酸結合ドメインを持つ(Monia, B.P. 他(1989) J. Biol. Chem. 264:4093-4103)。この融合タンパク質は重要な中間体であり、細胞の翻訳活性とタンパク質分解活性との同時制御と、非活性酵素の、特異的細胞部位への局在とを仲介するようである。送達されると、C末端加水分解酵素がこの融合タンパク質を切断し、機能的Ubを放出する(Monia他、前出)。

40

【0017】

Ub抱合酵素(E2)は、種々のUCS経路における基質特異性のために重要である。全てのE2

50

は、UBCドメインと呼ばれるおよそ16kDaの保存されたドメインを有する。UBCドメインは全てのE2において少なくとも35%の同一性があり、ユビキチン-酵素チオールエステル形成に必要な、中心に位置するシステイン残基を含む(Jentsch、前出)。良く保存されたプロリンリッチエレメントが、N末端から活性システイン残基に位置する。この保存されたドメイン以外の構造的変異を用いて、E2酵素を分類する。クラスIのE2は、ほぼ例外なく、保存UBCドメインから成る。クラスIIのE2には、基質特異性および細胞局在に寄与する、多様な関連のないC末端延長がある。クラスIIIのE2には、酵素調節または基質特異性に関与すると思われる特有のN末端延長がある。

【0018】

有糸分裂サイクリン特異的E2 (E2-C)は、保存されたUBCドメイン、他のE2では見出されない30アミノ酸のN末端延長、この延長に隣接する7アミノ酸特異配列により特徴付けられる。サイクリンへのE2-Cの高親和性と共にこれらの特性はE2-Cを、E2の新しいクラスとして同定する(Aristarkhov, A.他(1996) Proc. Natl. Acad. Sci. 93:4294-99)。

【0019】

ユビキチン-タンパク質リガーゼ(E3)はユビキチン抱合プロセスの最終ステップ、基質とのユビキチンの共有結合を触媒する。E3は本プロセスの特異性の決定に重要な役割を果たす。現在のところ同定されているE3はわずかである。E3リガーゼの1つのタイプは、HECT(E6-APのC末端に相同的)ドメインタンパク質ファミリーである。HECTファミリーの1つのメンバーであるE6-AP (E6関連タンパク質)は、ヒト乳頭腫ウイルス(HPV) E6腫瘍性タンパク質と同様に、p53のユビキチン結合と分解のために必要である(Scheffner他(1993) Cell 75:495-505)。HECTタンパク質のC末端ドメインは、高度に保存されたユビキチン結合システイン残基を含む。様々なHECTタンパク質のN末端領域は多様であり、特異的な基質認識に関与すると思われる(Huibregtse, J.M他(1997) Proc. Natl. Acad. Sci. USA 94:3656-3661)。SCF (Skp1-Cdc53/Cullin-Fボックス受容体)ファミリーのタンパク質は、別のグループのユビキチンリガーゼを構成する(Deshaies, R. (1999) Annu. Rev. Dev. Biol. 15:435-467)。複数のタンパク質がSCF複合体内へ動員される。このタンパク質としては、Skp1、cullin(カリン)、および、或るFボックスドメイン含有タンパク質を含む。このFボックス蛋白質はユビキチン結合反応のために基質に結合する。また、基質特異性を決定し、基質を反応のためにその方向を位置付ける役割を果たす可能性がある。Skp1はFボックス蛋白質とカリンとの双方と相互作用する。また、Fボックス蛋白質とカリンとを、E2酵素からタンパク質基質へのユビキチンの転移のために、複合体内で位置づける過程に関わる可能性がある。SCFリガーゼ類の基質としては、CDK活性の調節、転写の活性化、シグナル伝達、動原体のアセンブリ、およびDNA複製に関わるタンパク質類を含む。

【0020】

Sgt1は、酵母での、Skp1における突然変異に起因する、動原体機能の欠損を抑止する遺伝子のスクリーニングにおいて同定された(Kitagawa, K. 他(1999) Mol. Cell 4:21-33)。Sgt1は、Skp1と相互作用する。また、SCFユビキチンリガーゼと会合する。Sgt1における欠損は、細胞周期のG1期またはG2期の何れかで細胞の停止を引き起こす。或る酵母Sgt1ヌル変異体は、ヒトSgt1によってレスキューされ得る。これは、多くの種にわたってSgt1機能が保存されていることを示す。Sgt1は、酵母での動原体複合体のアセンブリに必要である。

【0021】

UCSの異常な活性は、多数の疾患と障害とに関連付けられている。これらには例えば悪液質(Llovera, M. 他(1995) Int. J. Cancer 61: 138-141)、腫瘍抑制タンパク質p53の分解(Ciechanover, 前出)、および、アルツハイマー病に見られるような神経変性(Gregori, L. 他(1994) Biochem. Biophys. Res. Commun. 203: 1731-1738)を含む。ユビキチン抱合は抗原提示の律速段階なので、ユビキチン分解経路はまた、免疫応答において決定的な役割を持つ可能性がある(Grant E.P. 他(1995) J. Immunol. 155: 3750-3758)。

【0022】

数種の細胞増殖障害は、細胞周期中の進行を通常は制御するタンパク質複合体の変化によ

り同定され得る。或る初期治療ストラテジーには、細胞周期調節に関与するタンパク質群の操作により、細胞周期進行への制御を再確立することが関係している(Nigg, E.A. (1995) BioEssays 17:471-480)。

【0023】

腫瘍壊死因子 (TNF) と関連サイトカイン類とは、リンパ系細胞においてアポトーシスを誘発する(概説はNagata, S. (1997) Cell 88:355-365)。TNFの、その受容体への結合は或るシグナル伝達経路の引き金となり、その結果、カスパーゼと呼ばれる関連プロテアーゼ群のカスケードの活性化が生じる。このようなカスパーゼの1種であるICE(インターロイキン1 β 変換酵素)はシステインプロテアーゼであり、ICEは、ICE自己切断によって産生される、2種の大サブユニットと2種の小サブユニットとからなる(Dinarello, C. A. (1994) FASEB J. 8:1314-1325)。ICEは、主に単球で発現される。ICEはサイトカイン前駆体であるインターロイキン1 β を活性型へ変換し、その活性型が、急性炎症、慢性炎症、骨吸収、骨髄性白血病、および他の病理プロセスにおいて中心的役割を果たす。ICEと関連カスパーゼ類とは、形質移入した細胞株内で過剰発現すると、アポトーシスを起こす。

【0024】

アポトーシス性エフェクター経路の最終ステップは、核DNAの断片化である。最近、カスパーゼ活性をDNA断片化に連結する新規の因子が同定された(Xuesong, L. 他(1997) Cell 89:175-184)。この因子、DNA断片化因子45(DFF-45)は、カスパーゼによってタンパク質的に活性化される。また、DNA断片化に必要である。DFF-45は331アミノ酸の長さである。また、細胞内で、第二の特徴付けられていない因子とのヘテロ二量体として存在する。DFF-45のアミノ酸配列が示すところではDFF-45はヌクレアーゼではなく、このことは、DFF-45が下流ヌクレアーゼを活性化することを示唆する。また、DFF-45に関連する或るタンパク質をコードするmRNAが、マウス脂肪生成細胞から単離されている(Danesch, U. 他(1992) J. Biol. Chem. 267:7185-7193)。このmRNAの発現は、ステロイド処理した、分化中の脂肪細胞内で誘発される。予測されたタンパク質であるFSP-27(脂肪細胞特異的で27キロダルトン)は高度に塩基性であり、予測等電点は10である。

【0025】

アポトーシスの調節不全は最近、多くのヒト疾患の発生機序における有意な因子として認識されている。例えば、アポトーシス低下によって生じた過剰な細胞生存は、細胞増殖と免疫応答とに関する諸障害に寄与しうる。これらの障害としては、癌、自己免疫疾患、ウイルス感染、および炎症を含む。一方、アポトーシス増加によって生じた過剰な細胞死は、変性障害および免疫不全症(例えばAIDS、神経変性疾患、および骨髄異形成症候群)をもたらしうる(Thompson, C.B. (1995) Science 267:1456-1462)。

【0026】

胚形成

哺乳類の胚形成プロセスには、受精後の最初の数週間の発達を含む。この期間に、胚形成は、単一の受精卵から、3種の胚組織の形成へ、次に胚へと進む。この胚は、その殆どの内部器官と、その全ての外的特徴とを持つ。

【0027】

正常なコースの哺乳類胚形成は、多数の遺伝子と組織との、妥当な時間的および空間的な調節に依存する。これらの調節プロセスは、マウスで熱心に研究されている。依然として理解の乏しい或る必須プロセスが、受精後の胚ゲノムの活性化である。マウス卵母細胞が成長するにつれ、卵母細胞は転写物を蓄積する。転写物は、直接にタンパク質へ翻訳されるか、または調節されたポリアデニル化による後の活性化のために貯蔵されるかの何れかである。その後の減数分裂型成熟と排卵との間、母性ゲノムは転写的に不活性であり、殆どの母性転写物は、2細胞期における接合体遺伝子群の活性化に先立ち、または伴い、脱アデニル化および／または分解される(Stutz, A. 他(1998) Genes Dev. 12:2535-2548)。母性から胚性への移行には、卵母細胞転写物の分解を伴うが接合体転写物の分解は伴わず、また、胚ゲノムの活性化と、早期発達に順応するための細胞周期進行の誘導とを伴う。

【0028】

MATER (Maternal Antigen That Embryos Require: 胚が必要とする母性抗原)は初め、卵巣免疫を持つマウスに由来する抗体群の標的として同定された(Tong, Z-B., and Nelson, L.M. (1999) *Endocrinology* 140:3720-3726)。MATERをコードする遺伝子の発現は卵母細胞に限られ、この遺伝子を、限られた数の、哺乳類での既知の母性効果遺伝子の1つにしている(Tong, Z-B., 他(2000) *Mamm.Genome* 11:281-287)。MATERタンパク質は2細胞から先の胚の発達に必要である。これは、この遺伝子が非活性化されたマウスでの予備結果に基づく。1111アミノ酸のMATERタンパク質内には、或る親水性リピート領域(アミノ末端内)と、或る、14ロイシンリッチリピート群を持つ領域(カルボキシル末端内)とを含む。これらリピートは、タンパク質間相互作用に決定的である、ブタのリボヌクレアーゼ阻害剤に見られる配列に似ている。

10

【0029】

減数分裂型成熟と排卵との際の母性転写物の分解は、排卵直前のリボヌクレアーゼの活性化に関わる可能性がある。したがって、MATERの機能は、母性リボヌクレアーゼに結合し、接合体転写物の分解を防ぐことである可能性がある(Tong (2000) 前出)。卵母細胞発達と胚形成における役割に加え、MATERはまた、卵巣免疫の発生機序に関わる可能性がある。理由は、MATERが、自己免疫性卵巣炎を持つマウスにおいて自己抗体の標的であるからである(Tong (1999) 前出)。

【0030】

母性mRNA D7は、アフリカツメガエルでは中程度に豊富な転写物である。その発現は、卵形成と早期胚形成とで最高であり、おそらくこれらの時期に限られる。D7タンパク質は卵母細胞には存在せず、卵母細胞成熟時に初めて蓄積し始める。D7タンパク質レベルは胚の発生の初日に最高であり、その後は減少する。D7タンパク質の損失は、成熟プロセス自体に影響し、胚小胞分解の時間経過を有意に遅らせる。このように、D7は、卵母細胞成熟に関わる、新規に記載されたタンパク質である(Smith R.C., 他(1988) *Genes Dev.* 2(10):1296-306)。

20

【0031】

多くの他の遺伝子が、以後の胚形成の諸ステージに関わる。受精後、卵母細胞の誘導が各卵管の遠位端の卵管采によって成され、卵母細胞は卵管に入り、通過し、そして子宮へ誘導される。子宮内膜で諸変化が起こり、受精卵の着床と胚の発達とを支持する準備ができる。幾つかの段階での分割が起きた後、今や約100細胞を持つ胚盤胞となった分割卵が、子宮に入る。子宮に到達すると、発達中の胚盤胞は通常、子宮腔に更に2~4日とどまった後、子宮内膜(子宮の内層)内に着床する。着床は、栄養膜細胞の作用による。栄養膜細胞は、胚盤胞の表面を覆って発達する。栄養膜細胞は、タンパク分解酵素群を分泌し、これら酵素が、内膜の細胞を消化し液化する。侵入プロセスの概説は、Fisher および Damsky (1993; *Semin Cell Biol* 4:183-188)並びにGraham および Lala (1992; *Biochem Cell Biol* 70:867-874)にある。着床が起きると、栄養膜細胞および他の下層細胞は迅速に増殖し、胎盤および、妊娠の種々の膜を形成する(Guyton, A.C. (1991) *Textbook of Medical Physiology*, 8th ed. W.B. Saunders Company, Philadelphia 915-919ページを参照)。

30

【0032】

胎盤は、発達する胎児を保護し養う、必須の役割を持つ。大抵の種では、合胞体栄養細胞層が、胎盤の外側の、胎児-母体インターフェースに在る。これは連続構造で、1細胞の深さを持ち、構成物である栄養膜細胞の融合によって形成される。合胞体栄養細胞は、母体-胎児間交換に、胎児発生時の組織リモデリングに、また、発生時の胎児の、母性免疫応答からの保護において重要な役割を果たす(Stoye, J.P. および Coffin, J.M. (2000) *Nature* 403:715-717)。

40

【0033】

syncytin(シンシチン)と言う遺伝子は、ヒトの内因性の欠損性プロウイルスのエンベロップ遺伝子である。syncytinは、高レベルで胎盤に発現し、精巣ではより弱く発現するが、いかなる他の組織にも検出されない(Mi, S. 他(2000) *Nature* 403:785-789)。胎盤内のsyncytin発現は、合胞体栄養細胞に限定される。レトロウイルスenvタンパク質類はしばし

50

ば細胞融合イベント群の促進に関わるので、syncytinは、栄養膜細胞の、合胞体栄養細胞層への融合の調節に関わりうると考えられた。諸実験が示したところでは、syncytinは、細胞融合をin vitroで仲介しうる。また、抗syncytin抗体類は、胎盤の細胞性栄養細胞の融合を阻害しうる(Mi, 前出)。また、或る保存された免疫抑制ドメインが、レトロウイルスのエンベロープタンパク質群に存在し、syncytinのアミノ酸残基373-397に見られるが、このドメインは、発達する胚に対する母性免疫応答の防止に関わる可能性がある。

【0034】

syncytinはまた、栄養膜侵入性の調節に、栄養膜の融合と最終分化とを誘発することによって関わる可能性がある(Mi, 前出)。子宮壁の不十分な栄養膜浸潤は、胎盤障害(例えば子癇前症)または妊娠誘発高血圧を伴う。一方、無制御な栄養膜侵入は、絨毛癌およびその他の妊娠性の栄養膜疾患で観察される。このように、syncytin機能は、これら疾患に関わる可能性がある。

10

【0035】

細胞分裂

細胞分裂は、全ての生物がこれによって成長し生殖する、基本的プロセスである。単細胞生物(例えば酵母や細菌)では、各細胞分裂により生物の数が2倍となる。一方、多細胞の種では、消耗またはプログラム細胞死により消失した細胞を置換するために、また、新たな組織または器官を作製するための細胞分化のために、多数回の細胞分裂が必要とされる。細胞分裂周期の詳細は異なりうるが、基本的なプロセスは3つの主要な現象からなる。最初の現象である間期には、細胞分裂の準備、DNAの複製、必須タンパク質の作製が含まれる。2番目の現象である有糸分裂では、核物質が分裂して細胞の両側に分離する。最終の現象である細胞質分裂は、細胞質の分配と分裂である。細胞周期移行の順序およびタイミングは、種々のチェックポイント(検問点)で正または負の調節回路によりプロセスを制御する細胞周期調節システムの制御下にある。

20

【0036】

細胞周期の調節された進行は、成長制御経路群と、基礎的な細胞周期機構との統合に依存する。細胞周期調節因子群は、ヒトcDNAと酵母cDNAから、過剰発現時に酵母接合フェロモン経路の細胞周期停止シグナル群をブロックまたは活性化するcDNAを選択して同定されている。既知の調節因子としては、ヒトCPR(細胞周期進行回復)遺伝子群(例えばCPR8およびCPR2)、並びに酵母CDC(細胞分裂制御)遺伝子群(CDC91を含み、停止シグナル群をブロック)を含む。CPR遺伝子群は、種々のタンパク質(例えばサイクリン、腫瘍抑制因子結合タンパク質、シャペロン、転写因子、翻訳因子、およびRNA結合タンパク質を含む)を発現する(Edwards, M.C. 他(1997) Genetics 147:1063-1076)。

30

【0037】

幾つかの細胞周期移行(細胞が有糸分裂に入ったり出たりすることを含む)は、サイクリン依存性キナーゼ(Cdk)類の活性化および阻害に依存する。Cdk類の構成は、キナーゼサブユニットであるCdkと、活性化サブユニットであるサイクリンとから成り、これらの複合体は、多くのレベルの調節を受ける。単一のCdkが出芽酵母(*Saccharomyces cerevisiae*)と分裂酵母(*Saccharomyces pombe*)に在るようである。一方、哺乳類は、種々の特化したCdkを持つ。サイクリンはサイクリン依存性プロテインキナーゼ(Cdk)類へ結合し活性化して作用する。Cdkは次に、有糸分裂プロセスに関与する選択されたタンパク質をリン酸化し活性化する。Cdk-サイクリン複合体は、リン酸化によって、また、CDC4およびCDC53等の分子の関わる、標的を定めた分解によって、正負両方に調節される。また、Cdk類は更に、阻害剤その他のタンパク質(例えばSuc1、これはCdkの、調節因子群への特異性またはアクセス可能性を修正)への結合によって調節される(Patra, D. および W.G. Dunphy (1996) Genes Dev.10:1503-1515; 並びに Mathias, N. 他(1996) Mol.Cell Biol. 16:6634-6643)。

40

【0038】

生殖

男女の生殖系は複雑であり、また、成長と発達との多くの側面に関わる。男女生殖系の解

50

剖学と生理学との概説は (Guyton, A.C. (1991) Textbook of Medical Physiology, W.B. Saunders Co., Philadelphia PA, 899-928ページ) にある。

【0039】

男性生殖系としては精子形成のプロセスを含み、ここで精子が形成される。また男性生殖機能の調節は、種々のホルモンと、それらの、付索性器、細胞代謝、成長、および他の身体機能への効果によって成される。

【0040】

精子形成は思春期に、下垂体前葉から放出される性腺刺激ホルモンによる刺激の結果として始まる。未熟な精子(精原細胞)は数回の有糸細胞分裂を経た後、減数分裂と完全な成熟を経験する。精巣は数種の男性ホルモンを分泌する。最も豊富なものはテストステロンであり、これは、未熟な精子の成長と分裂とに、また、男性の身体の男性的特徴に必須である。3種の他の男性ホルモン、性腺刺激ホルモン放出ホルモン(GnRH)、黄体形成ホルモン(LH)、および卵胞刺激ホルモン(FSH)が、性機能を制御する。

10

【0041】

子宮、卵巣、卵管、膈、および乳房が、女性生殖系を構成する。卵巣と子宮とは各々、卵子の源泉であり、胎児発達の部位である。卵管と膈とは付属器官であり、それぞれ子宮の上と下に付いている。子宮と卵巣との両方が、女性の生涯における生殖能力の発生と喪失で、更なる役割を有する。乳房の主要な役割は泌乳である。卵巣、子宮、下垂体、視床下部、副腎、その他の組織からの多数の内分泌シグナルが、生殖と泌乳とを協調させる。これらシグナルは毎月の月経周期の期間中および女性の生涯の期間中に変動する。同様に、これら内分泌シグナルへの、生殖器の感度は、女性の生涯の期間中に変動する。

20

【0042】

卵巣、下垂体、および視床下部への、正負のフィードバックの組合せが、月ごとの排卵および子宮内膜周期での生理変化を制御する。下垂体前葉は、2種の主要な性腺刺激ホルモンである、卵胞刺激ホルモン(FSH)と黄体形成ホルモン(LH)とを分泌する。これは、ステロイド類の負のフィードバックによって(最も著しくは卵巣エストラジオールによって)調節される。受精が起きない場合、エストロゲンとプロゲステロンとのレベルが低下する。この、卵巣ホルモンの急激な減少は、月経(内膜の剥離)を生じる。

【0043】

ホルモンは更に、妊娠、分娩、泌乳、および閉経の、全ステップを支配する。妊娠時には、多量のヒト絨毛性ゴナドトロピン(hCG)、エストロゲン、プロゲステロン、およびヒト絨毛性ソマトマンモトロピン(hCS)が、胎盤によって形成される。hCGは黄体形成ホルモンに似た糖タンパク質であり、黄体が、更なるプロゲステロンとエストロゲンとを産生し続け、卵子が受精しないときに起きるように萎縮することの無い様に刺激する。hCSは成長ホルモンに似ており、胎児の栄養摂取に死活的である。

30

【0044】

女性乳房はまた、妊娠時に成熟する。多量のエストロゲンが胎盤から分泌され、乳管系の成長と分枝形成との引き金となる。一方、泌乳の開始は、下垂体によるプロラクチンの分泌によって成される。

【0045】

分娩には、幾つかのホルモンの変化を伴う。これらの変化が、子宮の収縮性を、妊娠の終了に向けて次のように増大させる。エストロゲンのレベルが、プロゲステロンのレベル以上に増加する。オキシトシンが、神経下垂体によって分泌される。同時に、オキシトシンへの子宮の感度が増加する。胎児自身がオキシトシン、コルチゾル(副腎から)、およびプロスタグランジンを分泌する。

40

【0046】

閉経は、大半の卵胞が変性したときに起きる。卵巣はその後、より少ないエストラジオールを産生し、下垂体と視床下部での負のフィードバックを低下させる。循環するFSHとLHとの平均レベルが増加する(排卵周期が続くにも関わらず)。したがって、卵巣は性腺刺激ホルモンへの応答性がより低く、また、月経周期の間の時間が増加する。結果的に、月経

50

出血が止まり、生殖能力が終了する。

【0047】

細胞分化および増殖

組織成長には、複雑で秩序あるパターンの、細胞増殖、細胞分化、およびアポトーシスが関わる。細胞増殖は、細胞の数と、細胞の空間的構成との双方を維持するために調節される必要がある。この調節は、成長因子や他の有糸分裂促進因子(マイトジェン)などの細胞外信号と、DNA損傷または栄養飢餓などの細胞内合図とに応じて細胞周期の進行を制御するタンパク質の、適切な発現に依存する。細胞周期の進行を直接または間接的に調節する分子は、成長因子とその受容体、セカンドメッセンジャーとシグナル伝達タンパク質、発癌遺伝子産物、腫瘍抑制タンパク質、および有糸分裂促進因子などの、複数のカテゴリーに分類される。 10

【0048】

成長因子は初め、細胞増殖の促進に必要な血清因子として記載された。殆どの成長因子は大きな分泌ポリペプチドであり、その局所環境にある細胞に作用する。成長因子は、特異的な細胞表面受容体に結合して活性化させ、細胞内シグナル伝達カスケードを開始する。多くの成長因子受容体は、リガンド結合時に自己リン酸化を受ける受容体チロシンキナーゼとして分類される。自己リン酸化は、受容体が、SH2ドメインまたはSH3ドメイン(Src相同性領域2又は3)の存在で特徴付けられるシグナル伝達タンパク質と相互作用することを可能にする。これらタンパク質は次に、小さなGタンパク質(例えばRas、Rab、およびRho)の活性状態の調整を、GTPアーゼ活性化タンパク質(GAP)、グアニンヌクレオチド放出タンパク質(GNRP)、および他のグアニンヌクレオチド交換因子と共に行う。小さなGタンパク質は、他の下流イベント(例えばマイトジェン活性化プロテインキナーゼ(MAPキナーゼ)のカスケード)を活性化する分子スイッチとして作用する。MAPキナーゼは最終的に、有糸分裂促進遺伝子の転写を活性化する。 20

【0049】

成長因子に加え、小さなシグナル伝達ペプチドとホルモンも、細胞増殖に影響する。これら分子は主に他のクラスの受容体、即ち三量体のGタンパク質共役受容体(GPCR)に結合する。GPCRは、免疫細胞、神経細胞、および神経内分泌細胞の表面に主に見られる。リガンド結合時に、GPCRは三量体Gタンパク質を活性化し、Gタンパク質は次に、細胞内セカンドメッセンジャー(例えばホスホリパーゼC、Ca²⁺、およびサイクリックAMP)のレベル増加の引き金となる。大抵のGPCR仲介シグナル伝達経路は、直接の分裂促進効果を持つ、他のシグナリング分子の分泌または分解を起こして間接的に細胞増殖を促進する。これらのシグナル伝達カスケードはしばしば、キナーゼとホスファターゼとの活性化を伴う。数種の成長因子(例えばトランスフォーミング成長因子 β (TGF- β)ファミリーの幾つかのメンバー)は、数種の細胞上では細胞増殖を刺激するよう作用し、他の細胞では増殖を阻害するよう作用する。成長因子はまた、或る細胞を或る濃度では刺激し、同じ細胞を他の濃度では阻害することがある。大抵の成長因子はまた、細胞成長と細胞分裂との調節以外に、多数の他の作用を持つ。それら因子は、状況に応じ、細胞の増殖、生存、分化、遊走、または機能を制御できる。例えば腫瘍壊死因子/神経成長因子(TNF/NGF)ファミリーは、細胞死を活性化または阻害でき、また、増殖と分化とを調節できる。細胞応答は、細胞のタイプ、細胞の分化のステージ、トランスフォーメーション状態、どの表面受容体が刺激されるか、および、細胞に作用する刺激のタイプに依存する(Smith, A. 他(1994) Cell 76:959-962; およびNocentini, G. 他(1997) Proc. Natl. Acad. Sci. USA 94:6216-6221)。 30 40

【0050】

組織内で隣接する細胞群は、成長因子について競合する。また、灌流した系内で「無限」量を提供されると、細胞分裂の密度依存性阻害に到達するまでに、一層高い細胞密度まで成長することとなる。細胞はしばしば、細胞分裂の足場依存性をも示す。この足場依存性は、細胞骨格を細胞外マトリクス(ECM)と連結する接着点の形成に関連する可能性がある。ECM成分の発現は、成長因子によって刺激されうる。例えばTGF- β は、線維芽細胞を刺 50

激し、種々のECMタンパク質(これにはフィブロネクチン、コラーゲン、およびテネイシンを含む)を産生させる(Pearson, C.A. 他(1988) EMBO J. 7:2677-2981)。事実、数種の細胞タイプに特異的なECM分子(例えばラミニンまたはフィブロネクチン)は、成長因子として作用しうる。テネイシンCとテネイシンRとは発生期と罹患時との神経組織に発現され、刺激性／抗接着特性または阻害特性を各々、軸索成長に提供する(Faissner, A. (1997) Cell Tissue Res. 290:331-341)。

【0051】

癌は、正常な細胞遺伝子に由来する発癌遺伝子類の活性化を伴う。これら発癌遺伝子は腫瘍性タンパク質をコードし、このタンパク質が正常な細胞を悪性細胞に転換する。数種の腫瘍性タンパク質は正常タンパク質の突然変異アイソフォームであり、他の腫瘍性タンパク質は、発現の部位または量に関して異常に発現する。後者のカテゴリーの腫瘍性タンパク質は、細胞増殖の転写調節を変更することにより癌の原因となる。5クラスの腫瘍性タンパク質が、細胞周期制御に影響することが既知である。これらクラスには、成長因子、成長因子受容体、細胞内シグナル伝達因子、核内転写因子、および、細胞周期制御タンパク質を含む。ウイルス発癌遺伝子は、ヒト細胞が数種のウイルスに感染するとヒトゲノムに組込まれる。ウイルス発癌遺伝子の例としては、v-src、v-abl、およびv-fpsを含む。腫瘍と転移とに関わるタンパク質の過剰発現に関連する多くの症例が報告されている。Mtal遺伝子は、マウスでは、転移性腫瘍を示す細胞株と組織との双方においてクローン化されている(Simpson, A. 他(2001) Gene 273:29-39)。メラノーマ抗原コード遺伝子(MAGE)ファミリーのタンパク質の発現も、多くの腫瘍で検出されている。GAC1はロイシンリッチリピートスーパーファミリーの新規メンバーであり、悪性グリオーマでは増幅され過剰発現する(Almeida, A. 他(1998) Oncogene 16: 2997-3002)。

【0052】

多くの発癌遺伝子が同定され特徴付けられている。発癌遺伝子としてはerbA、erbB、her-2、突然変異G_s、src、abl、ras、crk、jun、fos、myc、および突然変異した腫瘍抑制遺伝子(例えばRB、p53、mdm2、Cip1、p16、およびサイクリンD)を含む。正常遺伝子から発癌遺伝子への転換は、染色体転座によっても起こりうる。フィラデルフィア染色体は、慢性骨髄性白血病と、或るサブセットの急性リンパ芽球性白血病との特徴であり、これは、9番染色体と22番染色体との間の相互転座の結果である。この転座は、プロトオンコジーンc-ablの或る切り詰められた部分を、22番染色体のブレイクポイント(切断点)クラスター領域(bcr)へ移動させる。

【0053】

腫瘍抑制遺伝子は、細胞増殖の調節に関係する。腫瘍抑制遺伝子の機能を低減または喪失させる突然変異の結果、無制御な細胞増殖が起こる。例えば網膜芽腫遺伝子産物(RB)は、非リン酸化状態では、数種の早期応答遺伝子と結合し、それらの転写を抑制して、細胞分裂をブロックする。RBのリン酸化は、RBをこれら遺伝子から解離させ、抑制を解除し、細胞分裂を進行させる。

【0054】

SEB (SET結合タンパク質) は新規な核タンパク質である。SEBは酵母2ハイブリッド系とヒト細胞内でSETと相互作用する。SETは転座切断点コード蛋白質であり、急性未分化白血病に見られる。SEBはまた、1つの腫瘍性タンパク質Ski相同領域、6つのPEST配列群、および3つの連続PPLPPPPリピート群を、C末端に持つ。SEB mRNAの発現は、全ての試験されたヒト成人組織および細胞に遍在的に見られる。SETは、染色体18q21.1にマップされている。この領域にはまた、癌と白血病とでの欠失に関わる腫瘍抑制遺伝子群を含む(Minakuchi, M. 他(2001) Eru. J. Biochem. 268:1340-1351)。

【0055】

細胞分化

多細胞生物は、各細胞は互いに、その遺伝的資性において類似しているにもかかわらず、構造及び機能が著しく異なる多様な細胞型から構成される。細胞分化のプロセスによって、細胞群は、その構造と生理機能とが異なるようになる。多細胞生物の細胞は全て、単細胞

10

20

30

40

50

胞接合体の有糸分裂に起因する。接合体は全能性である。即ち、成体の如何なるタイプの細胞をも生じる能力を持つ。発生中に、接合体の細胞子孫はその全能性を失い、決定されて行く。細胞の予定運命が樹立すると、細胞は分化したと言われる。この細胞の全ての子孫は、同じタイプとなる。

【0056】

ヒトの成長と発達には、細胞増殖、および調節された細胞死と共に、細胞分化の空間的および時間的な調節を要する。これらのプロセスは協調し、生殖、老化、胚形成、形態形成、器官形成、ならびに組織の修復および維持を制御する。細胞分化に関わる諸プロセスはまた、病態（癌など）にも関与する。この場合、正常な細胞分化を調節する因子類が変容しており、癌細胞を未分化または低分化状態で増殖させる。

10

【0057】

分化の機序には転写と翻訳との細胞特異的な調節を伴うので、異なる遺伝子群が、異なる時期に異なる細胞で選択的に発現する。遺伝実験にショウジョウバエ (*Drosophila melanogaster*) を用いて、発生および分化時のパターン形成を制御する転写因子の、調節されたカスケードが同定されている。これらカスケードには、ホメオティック遺伝子群を含む。これら遺伝子は、ホメオボックスモチーフを持つ転写因子群をコードする。ホメオティック遺伝子の産物は、如何にしてこの昆虫の成虫原基が未分化細胞の塊から複雑な器官を持つ特異的な体節へ発達するかを決定する。ショウジョウバエの細胞分化と発達とに関わることが見出された多くの遺伝子は、相同体を哺乳類に有する。数種のヒト遺伝子は、そのショウジョウバエ相同体と等価の発達の役割を持つ。ショウジョウバエ *eyes absent* 遺伝子 (*eya*) のヒト相同体は *branchio-oto-renal* (鰓弓耳腎) 症候群の根底にある。この発達障害は、耳と腎臓とを冒す (Abdelhak, S. 他 (1997) *Nat. Genet.* 15:157-164)。ショウジョウバエ *slit* 遺伝子は、分泌性のロイシンリッチリピート含有タンパク質をコードする。このタンパク質は、正中線膠細胞によって発現される。また、正常な神経発生に必要である。

20

【0058】

細胞レベルでは、成長と発達との支配は、細胞が細胞周期に入ったり出たりする決定に、また、細胞の、最終分化状態へのコミットメントによって成される。細胞内の分化遺伝子発現は、細胞外シグナルとその他の環境合図へ応答して誘発される。このようなシグナルとしては、成長因子および他の分裂促進因子 (例えばレチノイン酸)、細胞間接触および細胞-マトリクス接触、並びに、環境因子 (例えば栄養シグナル、毒性物質、および熱ショック) を含む。分化の役割を果たしうる候補遺伝子群の同定は、*in vitro* での細胞分化の誘発時の、変容した発現パターンによって成されうる。

30

【0059】

細胞分化の最終ステップの結果、特化が生じる。特化の特徴は、特定タンパク質の産生 (例えば筋細胞での収縮タンパク、肝細胞での血清タンパク、および、赤血球前駆体でのグロビン) である。これら特化タンパクの発現は、少なくとも部分的に、細胞特異的転写因子群に依存する。例えばホメオボックス (*homobox*) 含有転写因子 PAX-6 は、脊椎動物の、早期の眼の決定、眼組織の特異化、および、正常な眼の発達に必須である。

【0060】

表皮分化の場合、分化特異的遺伝子群の誘発は、成長停止と共にまたはその後の何れかに起こる。また、不可逆的な成長停止を制御する分子イベント群に関連すると思われる。不可逆的な成長停止という早期イベントは、細胞が、皮膚の基底層から、最深部の *suprabasal layer* (基底上層) へ移行し、扁平上皮特異的遺伝子群を発現し始めるときに起きる。これら遺伝子としては、架橋したエンベロープの形成に関わる遺伝子 (例えばトランスグルタミナーゼ I および III、*involucrin* (インボルクリン)、*loricin* (ロリクリン)、および、スモールプロリンリッチリピート (SPRR) タンパク等) を含む。SPRR タンパク類は 8~10 kDa の分子量を持ち、プロリン、グルタミン、およびシステインが豊富である。また、類似した反復配列エレメントを持つ。SPRR タンパクは、強い二次構造を持つ構造タンパクであるか、または、金属結合タンパク (メタロチオネインなど) でありうる (Jetten, A. M. お

40

50

よび Harvat, B. L. (1997) J. Dermatol. 24:711-725; PRINTS Entry PR00021 PRORICH Small proline-rich protein signature)。

【0061】

Wnt遺伝子ファミリーの分泌シグナル分子は、真核細胞全体で高度に保存されている。Wntファミリーのメンバーは、軟骨テンプレート(cartilage template)内の軟骨細胞分化の調節に関わる。Wnt-5a、Wnt-5b、およびWnt-4遺伝子の発現は、ニワトリ四肢の軟骨形成領域で成される。Wnt-5aの発現は、軟骨膜(早期軟骨テンプレートのすぐ周辺の間葉細胞)で成される。Wnt-5a誤発現は、ニワトリ四肢で、軟骨細胞の成熟と、bone collar形成の開始を遅らせる(Hartmann, C. and Tabin, C.J. (2000) Development 127:3141-3159)。

【0062】

グリピカン(glypicans)は細胞表面へパラン硫酸プロテオグリカンの1ファミリーであり、これらは、細胞成長制御と分化とで重要な役割を果たす。cerebroglycanはへパラン硫酸プロテオグリカンであって神経系に発現される。これは、発達するニューロンの運動性の挙動に関わる(Stipp, C.S. 他(1994) J. Cell Biol. 124:149-160)。

【0063】

Notchは膠細胞の分化で活性な役割を果たす。また、神経突起の長さと構成とに影響する(概説は、Frisen, J. および Lendahl, U. (2001) Bioessays 23:3-7を参照)。Notch受容体シグナル経路は、多細胞種での多くの器官および組織の、形態形成と発達とに重要である。ショウジョウバエfringeタンパクは、翅の成虫原基の背腹境界でのNotchシグナル伝達経路の活性化を調整する。哺乳類fringe関連ファミリーメンバーは、分節時の境界決定

10

20

【0064】

最近、数種のタンパク質に、或る保存されたシステインリッチドメインがあることがわかり、これは約60アミノ酸残基からなり、LIMドメインと呼ばれる(Lin-11、Isl-1、Mec-3に由来)(Freyd G. 他(1990) Nature 344:876-879; Baltz R. 他(1992) Plant Cell 4:1465-1466)。LIMドメインには、7つの保存されたシステイン残基と、1つのヒスチジンとがある。LIMドメインは、2個の亜鉛イオンと結合する(Michelsen J.W. 他(1993) Proc. Natl. Acad. Sci. U.S.A. 90:4404-4408)。LIMはDNAとは結合せず、むしろ、タンパク間相互作用のインターフェースとして作用するようである。

【0065】

アポトーシス

多細胞生物の正常な発生、成長、および恒常性には、全身の組織での細胞の産生と破壊との間の慎重なバランスを要する。細胞分裂は、慎重に調整されるプロセスであり、多数のチェックポイントと制御機序とを持つ。これらの機序の設計は、DNA複製を調節し、不適切または過剰な細胞増殖を防ぐように成される。一方、アポトーシスは、遺伝的に制御されたプロセスであり、これにより、不要なまたは欠陥のある細胞が、プログラム細胞死を迎える。壊死細胞または傷害細胞と異なり、アポトーシス細胞は、その潜在的に損傷性の内容物を周辺組織に漏らしたり、炎症応答の引き金となったりすることなく、隣接細胞またはマクロファージに迅速に貪食される。

【0066】

アポトーシスは遺伝的に制御されたプロセスであり、これにより、不要なまたは欠陥のある細胞が、プログラム細胞死を迎える。細胞の選択的除去は、形態形成と組織リモデリングにとって、細胞増殖および分化と同じく重要である。アポトーシス不足の結果、過形成その他の、増加した細胞増殖に伴う障害を生じうる。アポトーシスはまた、免疫応答の決定的要素である。免疫細胞(例えば細胞障害性T細胞およびナチュラルキラー細胞)は、腫瘍細胞やウイルス感染細胞にアポトーシスを誘発して、疾患が広がるのを防ぐ。また、自己分子と外来分子とを区別しない免疫細胞は、アポトーシスで除去され自己免疫応答を避ける必要がある。

【0067】

アポトーシス細胞は、明確な形態変化を経る。アポトーシスの特質としては、細胞収縮、

30

40

50

核と細胞質との凝縮、および、形質膜トポロジーの変容を含む。生化学的には、アポトーシス細胞の特徴は、増加した細胞内カルシウム濃度、染色体DNAの断片化、および、新規な細胞表面成分の発現である。

【0068】

アポトーシスの分子機序は高度に保存されており、アポトーシスの鍵となる多くのタンパク調節因子およびエフェクターが同定されている。アポトーシスは一般に、細胞内伝達され結果として変容したパターンの遺伝子発現とタンパク活性とを生じるシグナルに応じて進行する。シグナル分子(例えばホルモンやサイトカイン)は、アポトーシスの刺激と阻害との双方を、細胞表面受容体との相互作用を介して行うことが知られる。転写因子もまた、アポトーシスの開始において重要な役割を果たす。多数の下流エフェクター分子(特にプロテアーゼ)が、細胞成分の分解と、他のアポトーシスエフェクター群のタンパク分解活性化とに関連付けられている。

10

【0069】

Bcl-2ファミリーのタンパク質は、他の細胞質タンパクと同様、アポトーシスの鍵となる調節因子である。少なくとも15種のBcl-2ファミリーメンバーが、3種のサブファミリーに存在する。Bcl-2タンパクは、哺乳類細胞とウイルスとで同定されている。また、各々が、4種のBcl-2相同性ドメイン(BH1~BH4)の少なくとも1つを持つ。これらドメインは高度に保存される。Bcl-2ファミリー蛋白にはBH1およびBH2ドメインが含まれ、これらはプロサバイバル(生存促進)サブファミリーのメンバーに見られるが、Bcl-2に最も似たタンパク類は4種すべての保存ドメインを持ち、種々の細胞障害性の挑戦に遭遇した際のアポトーシスの阻害を可能にしている。プロサバイバルサブファミリーのメンバーとしては、Bcl-2、Bcl-x_L、Bcl-w、Mcl-1、およびA1(哺乳類); NF-13 (ニワトリ); CED-9 (線虫 *Caenorhabditis elegans*); 並びにウイルス蛋白であるBHRF1、LMW5-HL、ORF16、KS-Bcl-2、およびE1B-19Kを含む。BH3ドメインは、プロアポトーシスサブファミリー蛋白の機能に必須である。2種のプロアポトーシスサブファミリーである、BaxとBH3とは各々、Bax、Bak、およびBok(別名Mtd)と、Bik、Blk、Hrk、BNIP3、Bim_L、Bad、Bid、およびEgl-1 (*C. elegans*)とを含む。BaxサブファミリーのメンバーはBH1、BH2、およびBH3ドメインを持ち、Bcl-2にかなり良く似ている。一方、BH3サブファミリーのメンバーは9~16残基のBH3ドメインのみを持つ以外は如何なる既知タンパクとも無関係であり、BikとBlkとのみが配列類似性を共有する。2種のプロアポトーシスサブファミリーのタンパク質は、プロサバイバルサブファミリー蛋白のアンタゴニストの可能性もある。この例証は *C. elegans* にあり、ここではEgl-1(アポトーシスに必要)が、CED-9に結合し、これを介して作用する(概説はAdams, J.M. および Cory, S. (1998) Science 281:1322-1326を参照)。

20

30

【0070】

プロアポトーシスサブファミリー蛋白および抗アポトーシスサブファミリー蛋白の間のヘテロ二量体化は、これらタンパク質サブファミリーの機能ヘタイトレーション的效果を持つようであり、これが示唆するところでは、各サブファミリーのメンバーの相対濃度が作用してアポトーシスを調節する。ヘテロ二量体化はプロサバイバル蛋白には不要だが、BH3サブファミリーには必須であり、Baxサブファミリーにはさほどでない。

【0071】

Bcl-2タンパクは2種のアイソフォーム、 α と β とを持ち、これらは選択的スプライシングによって形成される。Bcl-2はホモ二量体を形成し、また、BaxおよびBakタンパク、並びにBcl-XアイソフォームであるBcl-xとのヘテロ二量体を形成する。Baxとのヘテロ二量体化には無傷のBH1およびBH2ドメインを要する。またこれは、プロサバイバル活性に必要である。BH4ドメインも、プロサバイバル活性に関わるようである。Bcl-2はミトコンドリアの内膜と外膜に局在し、また核エンベロープと小胞体内にも局在し、発現は種々の組織で成される。Bcl-2のろ胞性リンパ腫(タイプII慢性リンパ性白血病)での関与が、染色体転座T(14;18) (q32;q21)に見られ、Bcl-2の関与は免疫グロブリン遺伝子領域に関わる。

40

【0072】

Bcl-xタンパク質は、アポトーシス細胞死の主要な調節因子である。選択的スプライシン

50

グの結果、3種のアイソフォームである、Bcl-xB、長いアイソフォーム、および短いアイソフォームを生じる。長いアイソフォームは細胞死抑制作用を示し、短いアイソフォームはアポトーシスを促進する。Bcl-xLはBaxおよびBakとのヘテロ二量体を形成するが、Baxとのヘテロ二量体化は、プロサバイバル(抗アポトーシス)活性に必要とは見えない。Bcl-xSはBcl-2とのヘテロ二量体を形成する。Bcl-xは、ミトコンドリア膜と、核周囲エンベロープとに見られる。Bcl-xSの高レベルでの発現は、発生中のリンパ球その他の、高速の代謝回転を受けている細胞に見られる。Bcl-xLは、成人の脳に、また、他の組織の、寿命の長い有糸分裂後細胞に見られる。Bcl-2と同様、BH1、BH2およびBH4ドメインは、プロサバイバル活性に関わる。

【0073】

Bcl-wタンパクは、殆ど全ての骨髄性細胞系の細胞質に、また多数の組織に見られる。最高の発現レベルは、脳、結腸、および唾液腺に見られる。Bcl-wタンパクは、低レベルでは精巣、肝臓、心臓、胃、骨格筋、および胎盤、並びに少数のリンパ性細胞系において発現される。Bcl-wは、BH1、BH2、およびBH4ドメインを持ち、この全ては、その細胞生存促進活性に必要である。Bcl-w遺伝子機能が相同組換えで破壊されたマウスは生存可能で、健康で、外見が正常であり、成体雌は正常な生殖機能を有したが、成体雄は不妊性であった。これら雄では、精子形成の最初の前思春期段階の大部分は無影響で、精巣は正常に発達した。しかし精細管が無秩序で、多数のアポトーシス細胞を有し、成熟精子を作れなかった。このマウスモデルは、数種のヒト男性不妊に適用できる可能性があり、この示唆するところでは、精巣のプログラム細胞死の改変は、生殖力の調整に有用であり得る(Print 20, C.G. 他(1998) Proc. Natl. Acad. Sci. USA 95:12424-12431)。

【0074】

ラット虚血脳の研究では、Bcl-wが、非虚血脳で正常なレベル(恒常的に低い)で発現されているのに比して、虚血脳で過剰発現されていた。更に、in vitro試験でBcl-wの作用機序を実験したところ、単離したラット脳ミトコンドリアは、Bcl-wも存在すると、組換えBaxまたは高濃度のカルシウムの添加に応答不能であった。正常な応答であれば、チトクロムcがミトコンドリアから放出されたであろう。また、組換えBcl-wタンパクは、カルシウム誘発によるミトコンドリア膜電位の喪失を阻害することが見られた。これは、透過性の変化を示す。これらの所見が共に示唆するところでは、Bcl-wは、虚血性神経細胞死に対する神経保護作用を示し、この保護を、ミトコンドリア死亡調節経路を介して達成する(Yan, C. 他(2000) J. Cereb. Blood Flow Metab. 20:620-630)。

【0075】

bfl-1遺伝子は、Bcl-2ファミリーの更なるメンバーであり、これもアポトーシスの抑制因子である。Bfl-1タンパクは175アミノ酸を持ち、また、BH1、BH2、およびBH3保存ドメインを持つ。これらドメインはBcl-2ファミリーメンバーに見られる。Bfl-1はまた、GlnリッチなNH2末端領域を持ち、NHドメイン1を欠く。これは他のBcl-2ファミリーメンバーと異なる。マウスA1タンパクは、Bfl-1と高い配列相同性を共有し、Bfl-1に見られる3種の保存ドメインを持つ。p53腫瘍抑制タンパクに誘発されるアポトーシスは、Bfl-1によって抑制される。これは、Bcl-2、Bcl-xL、およびEBV-BHRF1の作用に似る(D'Sa-Eipper, C. 他(1996) Cancer Res. 56:3879-3882)。Bfl-1は細胞内に見られる。最高発現は、造血区画(即ち血液、脾臓、および骨髄)に見られ、中等度の発現は肺、小腸、および精巣で、また最少発現は他の組織に見られる。Bfl-1はまた、血管平滑筋細胞および造血器悪性腫瘍(hematopoietic malignancies)に見られる。bfl-1の発現レベルと、胃癌の発症との間の相関が記述されており、その示唆するところでは、Bfl-1タンパクは、癌細胞生存の促進、または癌で胃癌の発症に関わる。(Choi, S.S. 他(1995) Oncogene 11: 1693-1698)。

【0076】

癌の特徴は、継続的で、または無制御な細胞増殖である。数種の癌には、正常なアポトーシス細胞死の抑制を伴う。治療のストラテジーには、細胞周期進行への制御を再確立することや、癌細胞のアポトーシスを選択的に刺激することを含みうる(Nigg, E.A. (1995) BioEssays 17:471-480)。癌の免疫防御としては、突然変異細胞での、腫瘍抑制因子による

10

20

30

40

50

アポトーシスの誘発と、腫瘍抗原の、Tリンパ球による認識とを含む。分裂促進ストレスへの応答は、しばしば転写のレベルで制御される。また、種々の転写因子によって調整される。例えばRel/NF- κ Bファミリーの脊椎動物転写因子は、放射線への炎症応答および免疫応答で中心的な役割を果たす。NF- κ Bファミリーとしては、p50、p52、RelA、RelB、cRel他のDNA結合タンパクを含む。p52タンパクはアポトーシスを誘発し、転写因子c-Junを上方制御し、c-Jun N末端キナーゼ1 (JNK1)を活性化する(Sun, L. 他(1998) Gene 208:157-166)。大抵のNF- κ Bタンパクは、DNA結合ホモ二量体またはヘテロ二量体を形成する。多くの転写因子の二量体化の仲介は、或る保存された配列が行う。bZIPドメインとして知られるその配列の特徴は、或る塩基性領域とそれに続く1つのロイシンジッパーである。Fas/Apo-1受容体(FAS)は、腫瘍壊死因子(TNF)受容体ファミリーのメンバーである。そのリガンド(Fasリガンド)と結合すると、膜貫通型のFASは、アポトーシスの誘発を、数種の細胞質タンパクを動員して行う。これらタンパクが死のシグナルを伝達する。FAS関連タンパク因子1 (FAF1)と呼ばれる、このようなタンパク質の1つが、マウスから単離された。また、L細胞でのFAF1の発現が、FAS誘発アポトーシスを増強することが示された(Chu, K. 他(1995) Proc. Natl. Acad. Sci. USA 92:11894-11898)。続いて、FAS関連因子群の単離が、多数の他の種(例えばショウジョウバエやウズラ)で成されている(Frohlich, T. 他(1998) J. Cell Sci. 111:2353-2363)。Fasからの死のシグナルの伝達で機能する別の細胞質タンパクがFas会合デスドメイン蛋白、略称FADDである。FADDは、死のシグナルを、FAS媒介およびTNFレセプター媒介アポトーシス経路の双方で、カスパーゼ8を活性化して伝達する(Bang, S. 他(2000) J. Biol. Chem. 275:36217-36222)。

10

20

【0077】

染色体DNAの断片化は、アポトーシスの特質の1つである。DNA断片化因子(DFF)タンパクは、2種のサブユニットからなる。40-kDaのカスパーゼ活性化ヌクレアーゼであるDFF40/CADと、その45-kDaの阻害因子であるDFF45/ICADである。DFF45/ICADの2種のマウス相合体であるCIDE-AおよびCIDE-Bが最近、報告されている(Inohara, N. 他(1998) EMBO J. 17:2526-2533)。哺乳類細胞でのCIDE-AとCIDE-Bとの発現はアポトーシスを活性化し、CIDE-A単独の発現は、DNA断片化を誘発した。また、FAS媒介アポトーシスは、CIDE-AとCIDE-Bとによって増進された。これは、これらタンパク質を、アポトーシスを媒介するエフェクターとして更に関連付ける。

【0078】

転写因子は、アポトーシスの開始において重要な役割を果たす。多数の下流エフェクター分子、特にプロテアーゼ(例えばカスパーゼと呼ばれるシステインプロテアーゼ)が、アポトーシスの開始段階と実行段階に関与している。カスパーゼの活性化は、プロサバイバルおよびプロアポトーシス型Bcl-2関連タンパクの競合作用による(Print, C.G. 他(1998) Proc. Natl. Acad. Sci. USA 95:12424-12431)。プロアポトーシス性シグナルは、タンパク質分解カスパーゼカスケードを引き起こすイニシエーターカスパーゼを活性化でき、標的タンパク質の加水分解と典型的なアポトーシス細胞死に導く。2つの活性部位残基、システインとヒスチジンは、触媒機序に関与すると考えられている。カスパーゼは最も特異的なエンドペプチダーゼの1つで、アスパラギン酸残基の後で切断する。

30

【0079】

カスパーゼは、不活性酵素前駆体として合成され、この不活性酵素前駆体は、小さなスペーサー領域によって分離される1つの大きな(p20)サブユニットと1つの小さな(p10)サブユニット、および、可変N末端プロドメインからなる。このプロドメインは、アポトーシスを高めたりまたは低めたりして影響しうる補助因子と相互作用する。活性化シグナルは、特異的なアスパラギン酸残基の自己タンパク質分解切断(カスパーゼ1の番号付け規則ではD297)およびスペーサーとプロドメインの除去を引き起こし、p10/p20ヘテロダイマーを残す。これらのヘテロダイマーの2つが、その小サブユニットを介して相互作用し、触媒的に活性な四量体を形成する。カスパーゼファミリーのいくつかのメンバーの長いプロドメインが、プロカスパーゼの二量体化と自己プロセッシングを促進していることが示されている。幾つかのカスパーゼはそのプロドメインに「デスエフェクタードメイン」を

40

50

含み、それによって、他のカスパーゼおよび、FADDタンパク質会合細胞死受容体 (death receptor) との自己活性化複合体に、あるいはTNF受容体複合体に、カスパーゼが補充されうる。さらに、カスパーゼファミリーの異なるメンバーからの2つの二量体が会合して、結果として得る四量体の基質特異性を変えることができる。

【0080】

アポトーシスの欠陥的調節は、アルツハイマー病でのニューロンの喪失に関わる。アルツハイマー病は進行性神経変性障害であり、特徴は、老人斑と、神経原線維変化との形成である。これらは、アミロイドβペプチドを含む。これらの斑は脳の辺縁皮質と連合皮質とに見られる。これには海馬、側頭皮質、帯状皮質、扁桃、基底核、および青斑核を含む。Bアミロイドペプチドは、アポトーシスを誘発しニューロンの死をもたらすシグナル経路 10
に関わる(Kajkowski, C. 他(2001) J. Biol. Chem. 276:18748-18756)。アルツハイマーの病理の早期には、生理変化が帯状皮質で見られる(Minoshima, S. 他(1997) Annals of Neurology 42:85-94)。進行したアルツハイマー病の患者では、蓄積する斑が、辺縁領域の神経構造を損傷し、最終的に、記憶プロセスをそこなう。

【0081】

腫瘍壊死因子 (TNF) と関連サイトカイン類とは、リンパ系細胞においてアポトーシスを誘発する(概説はNagata, S. (1997) Cell 88:355-365)。TNFの、その受容体への結合は、或るシグナル伝達経路の引き金となり、その結果、タンパク分解のカスパーゼカスケードの活性化が生じる。このようなカスパーゼの1種であるICE(インターロイキン1β変換酵素)はシステインプロテアーゼであり、ICEは、ICE自己切断によって産生される、2種の大サブユニットと2種の小サブユニットとからなる(Dinarello, C. A. (1994) FASEB J. 8:1314-1325)。ICEは、主に単球に発現する。ICEはサイトカイン前駆体であるインターロイキン1βを活性型へプロセッシングし、活性型が、急性炎症および慢性炎症、骨吸収、骨髄性白血病、および他の病理プロセスにおいて中心的役割を果たす。ICEと関連カスパーゼ類とは、形質移入した細胞株内で過剰発現すると、アポトーシスを起こす。 20

【0082】

カスパーゼ補充ドメイン(CARD)は、数種のアピカルカスパーゼ (apical caspases) のプロドメインに見られる。また、数種のアポトーシス調節分子、例えばApaf-2、RAIDD、および、アポトーシス蛋白の細胞性阻害物質(IAP)で保存される(Hofmann, K. 他(1997) Trends Biochem. Sci. 22:155-157)。アポトーシスでのCARDの調節的役割は、Apaf-1等のタンパク質をカスパーゼ9と会合させることでありうる(Li, P. 他(1997) Cell 91:479-489)。骨格筋と心筋との双方で発現する、CARD含有アポトーシス抑制因子 (ARC) をコードする或るヒトcDNAが、同定され特徴付けられている。ARCはアポトーシスの阻害因子として機能し、また、カスパーゼと選択的に相互作用する(Koseki, T. 他(1998) Proc. Natl. Acad. Sci. USA 95:5156-5160)。これら全ての相互作用は、アポトーシスの調整に明らかに影響する (Chan S.L.およびM.P. Mattson (1999) J. Neurosci. Res. 58:167-190; Salvenson, G.S. および V.M. Dixit (1999) Proc. Natl. Acad. Sci. USA 96:10964-10967に概説)。 30

【0083】

ES18は、アポトーシスの潜在的調節因子として、マウスT細胞で同定された(Park, E.J. 他(1999) Nuc. Acid. Res. 27:1524-1530)。ES18は428アミノ酸の長さであり、N末端プロリンリッチ領域、1つの酸性グルタミン酸リッチドメイン、および、1つの推定LXXLL核レセプター結合モチーフを持つ。ES18タンパクは、リンパ節と胸腺とで優先的に発現する。ES18発現のレベルは、T細胞胸腺腫S49.1では、デキサメタゾン、スタウロスポリン、またはC2セラミド (これらは、アポトーシスを誘発する) での治療に応じて増加する。ES18は、T細胞でのアポトーシス細胞死の刺激で役割を果たす可能性がある。 40

【0084】

ラット腹葉前立腺(RVP)は、ホルモン調節型アポトーシスの研究の或るモデル系である。RVP上皮細胞は、アンドロゲン剥奪に応答してアポトーシスを受ける。アポトーシスRVP内で上方制御されるメッセンジャーRNA(mRNA)転写物群が同定されている(Briehl, M. M. お 50

よび Miesfeld, R. L. (1991) Mol. Endocrinol. 5:1381-1388)。このような転写物の1つはRVP.1をコードする。アポトーシスでのこの正確な役割は決定されていない。RVP.1のヒト相同体であるhRVP1は、このラット蛋白に89%同一である(Katahira, J. 他(1997) J. Biol. Chem. 272:26652-26658)。hRVP1は220アミノ酸の長さであり、4つの膜貫通ドメインを持つ。hRVP1は、肺、腸、および肝臓で高発現される。興味深いことに、hRVP1は、ウェルシュ菌 (*Clostridium perfringens*) のエンテロトキシン(ヒト等の動物での下痢の原因物質)の低親和性レセプターとして機能する。

【0085】

サイトカイン媒介アポトーシスは、造血と免疫応答とにおいて重要な役割を果たす。骨髄性細胞は、マクロファージ、好中球、赤血球など血球の幹細胞前駆体であり、これは、特異的サイトカイン、例えば顆粒球/マクロファージ-コロニー刺激因子(GM-CSF)およびインターロイキン3(IL-3)などに応答して増殖する。GM-CSFまたはIL-3を剥奪すると、骨髄性細胞はアポトーシスを受ける。ネズミrequiem (req)遺伝子は、骨髄性細胞株FDCP-1での、このアポトーシス応答に必要な推定転写因子をコードする(Gabig, T. G. 他(1994) J. Biol. Chem. 269:29515-29519)。Reqタンパクは371アミノ酸の長さであり、1つの核局在シグナル、単一のKruppelタイプZnフィンガー、1つの酸性ドメイン、および、一群の、4種の独特なZnフィンガーモチーフ(金属結合に関わるシステインおよびヒスチジン残基が豊富)を持つ。reqの発現は骨髄性特異的でもアポトーシス特異的でもなく、これが示唆するところでは、付加的因子群が、骨髄性細胞アポトーシスでのReq活性を調節する。

【0086】

アポトーシスの調節不全は最近、多くのヒト疾患の発生機序における有意な因子として認識されている。例えば、アポトーシス低下によって生じた過剰な細胞生存は、細胞増殖と免疫応答とに関する諸障害に寄与しうる。これらの障害としては、癌、自己免疫疾患、ウイルス感染、および炎症を含む。一方、アポトーシス増加によって生じた過剰な細胞死は、変性障害および免疫不全症(例えばAIDS、神経変性疾患、および骨髄異形成症候群)をもたらす(Thompson, C.B. (1995) Science 267:1456-1462)。

【0087】

アポトーシスの調節不全は最近、多くのヒト疾患の発生機序における有意な因子として認識されている。例えば、アポトーシス低下によって生じた過剰な細胞生存は、細胞増殖と免疫応答とに関する諸障害に寄与しうる。これらの障害としては、癌、自己免疫疾患、ウイルス感染、および炎症を含む。一方、アポトーシス増加によって生じた過剰な細胞死は、変性障害および免疫不全症(例えばAIDS、神経変性疾患、および骨髄異形成症候群)をもたらす(Thompson, C.B. (1995) Science 267:1456-1462)。

【0088】

アポトーシスの欠陥的調節はまた、アルツハイマー病でのニューロンの喪失に関わる。アルツハイマー病は進行性神経変性障害であり、特徴は、老人斑と、神経原線維変化との形成である。これらは、アミロイドβペプチドを含む。これらの斑は脳の辺縁皮質と連合皮質とに見られる。これには海馬、側頭皮質、帯状皮質、扁桃、基底核、および青斑核を含む。βアミロイドペプチドは、アポトーシスを誘発しニューロンの死をもたらすシグナル経路に関わる(Kajkowski, C. 他(2001) J. Biol. Chem. 276:18748-18756)。アルツハイマーの病理の早期には、生理変化が帯状皮質で見られる(Minoshima, S. 他(1997) Annals of Neurology 42:85-94)。進行したアルツハイマー病の患者では、蓄積する斑が、辺縁領域の神経構造を損傷し、最終的に、記憶プロセスを不自由にする。

【0089】

癌

癌は、主要な公衆保健の関心事であり続けており、現行の予防策と治療とは、大半の患者の要求を満たしていない。癌(別称は新形成)の特徴は、継続的で無制御な細胞増殖である。癌は、3つのカテゴリーに分類できる。癌腫、肉腫、白血病である。癌腫は柔らかい上皮細胞の悪性成長であり、周辺組織を浸潤し、転移性腫瘍を生じうる。肉腫は、上皮起源のものや、結合組織に起因するものがある。白血病は造血組織の進行性悪性疾患であり、

10

20

30

40

50

特徴は、白血球とその前駆体との増殖であり、骨髄性(顆粒球または単球に由来)或いはリンパ球性(リンパ球に由来)に分類されうる。腫瘍化とは、腫瘍の発端からの成長の進行を指す。悪性細胞は、起源の組織内の正常な細胞に極めて類似のものや、未分化(低分化)なものがある。腫瘍細胞は、少数の核や、1個の大きな多形核を持ちうる。未分化細胞は、無秩序な塊に成長しうる。塊は血管化が乏しく、その結果、大きな領域が虚血壊死となっている。分化した新生細胞は、起源の組織と同じタンパク質を分泌しうる。癌は成長し、周辺組織に浸潤し、侵襲し、破壊する。これは、体腔または体表への直接播種によって、またはリンパ性転移や血行性転移によって成される。癌は、主要な公衆保健の関心事であり続けており、現行の予防策と治療とは、大半の患者の要求を満たしていない。腫瘍化の新形成プロセスの理解は、予後のおよび診断的重要性を持つ分子マーカー群の同定によって補助されうる。

10

【0090】

新形成プロセスの理解は、予後のおよび診断的重要性を持つ分子マーカー群の同定によって補助されうる。癌は、正常な細胞を悪性細胞へ転換しうる腫瘍性タンパク類を伴う。数種の腫瘍性タンパク質は正常タンパク質の突然変異アイソフォームであり、他の腫瘍性タンパク質は、発現の部位または量に関して異常に発現する。正常な細胞増殖は、成長因子の、細胞膜上のその受容体への結合で始まる。この結果、シグナル系が活性化し、これが核調節因子を誘導し活性化してDNA転写を開始させ、次に細胞分裂を生じる。細胞周期制御に影響する腫瘍性タンパクのクラスとしては、成長因子、成長因子受容体、細胞内シグナル伝達因子、核内転写因子、および、細胞周期制御タンパク質を含む。数種の癌特異的遺伝マーカー(例えば腫瘍抗原および腫瘍抑制因子)も同定されている。

20

【0091】

癌または悪性腫瘍は継続的な細胞増殖と細胞死によって特徴付けられる。また、3つのカテゴリーに分類しうる。癌腫、肉腫、白血病である。報告の示すところでは、約8人に1人の女性は乳癌にかかり、約10人に1人の男性(50歳以上)は前立腺癌にかかる(Helzlsouer, K. J. (1994) Curr. Opin. Oncol. 6:541-548; Harris, J. R. 他 (1992) N. Engl. J. Med. 327:319-328)。

【0092】

癌は、正常な細胞遺伝子に由来する発癌遺伝子類の活性化を伴う。これら発癌遺伝子は、腫瘍性タンパク質をコードし、このタンパク質が正常な細胞を悪性細胞に転換しうる。数種の腫瘍性タンパク質は正常タンパク質の突然変異アイソフォームであり、他の腫瘍性タンパク質は、発現の部位または量に関して異常に発現する。後者のカテゴリーの腫瘍性タンパク質は、細胞増殖の転写調節を変更することにより癌の原因となる。5クラスの腫瘍性タンパク質が、細胞周期制御に影響することが既知である。これらクラスには、成長因子、成長因子受容体、細胞内シグナル伝達因子、核内転写因子、および、細胞周期制御タンパク質を含む。場合によっては、発癌遺伝子は、レトロウイルスとDNAウイルスとによって活性化しうる。発癌遺伝子活性化は、ウイルスのゲノムの、宿主細胞のDNAへの組み込みの結果として起こる。これらの場合、感染した細胞を継続的に細胞分裂させる能力のある2つ以上の発癌遺伝子が活性化することもある。

30

【0093】

癌の特徴は、継続的で無制御な細胞増殖である。数種の癌には、正常なアポトーシス細胞死の抑制を伴う。新形成プロセスの理解は、予後のおよび診断的重要性を持つ分子マーカー群の同定によって補助されうる。癌は、正常な細胞を悪性細胞へ転換しうる腫瘍性タンパク類を伴う。数種の腫瘍性タンパク質は正常タンパク質の突然変異アイソフォームであり、他の腫瘍性タンパク質は、発現の部位または量に関して異常に発現する。正常な細胞増殖は、成長因子の、細胞膜上のその受容体への結合で始まる。この結果、シグナル系が活性化し、これが核調節因子を誘導し活性化してDNA転写を開始させ、次に細胞分裂を生じる。細胞周期制御に影響する腫瘍性タンパクのクラスとしては、成長因子、成長因子受容体、細胞内シグナル伝達因子、核内転写因子、および、細胞周期制御タンパク質を含む。数種の癌特異的遺伝マーカー(例えば腫瘍抗原および腫瘍抑制因子)も同定されている。

40

50

【0094】

癌治療の現行の形態としては、免疫抑制剤の使用を含む(Morisaki, T., Matsunaga H., 他(2000) *Anticancer Res.* 20: 3363-3373; Geoerger, B., Kerr, K., 他(2001) *Cancer Res.* 61:1527-1532)。細胞シグナリングに関わるタンパク質の同定は、また特に、免疫抑制剤の受容体として作用するタンパク質の同定は、抗腫瘍剤の開発を促進しうる。例えばイムノフィリンは、原核生物と真核生物との双方に見られる保存されたタンパク質の1ファミリーであり、免疫抑制剤に、種々の程度の特異性で結合する。このようなimmunophilicタンパク質の1グループが、ペプチジルプロリンシス-トランス異性化酵素(EC 5.2.1.8)ファミリー(PPIアーゼ、rotamase)である。これら酵素は、初めにブタ腎臓皮質から単離され、これらはタンパク折り畳みの加速を、オリゴペプチド内のプロリンイミドペプチド結合のシス-トランス異性化を触媒して行う(Fischer, G. および Schmid, F.X. (1990) *Biochemistry* 29: 2205-2212)。イムノフィリンファミリーには、シクロフィリン(例えばペプチジルプロリン異性化酵素A、即ちPPIA)とFK結合タンパク(即ちFKBP)とのサブファミリーを含む。シクロフィリンは多機能レセプター蛋白であり、これは、シグナル伝達活性(例えばシクロスポリンが媒介する伝達)に関わる。各ファミリーのPPIアーゼドメインは、種の間で高度に保存される。構造的には異なるが、これら多機能レセプター蛋白は、多数のシグナル伝達系路に関わる。また、折り畳みイベントと輸送イベントとに関連付けられている。

10

【0095】

イムノフィリン蛋白であるシクロフィリンは、免疫抑制剤であるシクロスポリンAと結合する。FKBPは別のイムノフィリンであり、FK506(またはラパマイシン)と結合する。ラパマイシンは免疫抑制剤であり、細胞を成長のG₁期で停止させ、アポトーシスを誘発する。シクロフィリン同様、このマクロライド抗生物質(産生は放線菌(*Streptomyces tsukubaensis*)に由来)の作用は、遍在性で主にサイトゾル性のイムノフィリン受容体に結合して成される。これらイムノフィリン/免疫抑制剤複合体(例えばシクロフィリンA/シクロスポリンA(CypA/CsA)、およびFKBP12/FK506)は、その治療成績の達成を、ホスファターゼであるカルシニューリン(カルシウム/カルモジュリン依存性タンパク質キナーゼであり、T細胞活性化に関与)を阻害して行う(Hamilton, G.S. および Steiner, J.P. (1998) *J. Med. Chem.* 41: 5119-5143)。ネズミfkbp51遺伝子の豊富な発現は、免疫組織(例えば胸腺およびTリンパ球)に見られる(Baughman, G., Wiederrecht, G.J., 他(1995) *Molec. Cell. Biol.* 15: 4395-4402)。FKBP12/ラパマイシン指示型免疫抑制は、TOR(酵母)またはFRAP(FKBP12-ラパマイシン会合タンパク、哺乳類細胞内)に結合して起きる。これらはラパマイシンのキナーゼ標的であり、正常な細胞成長パターンの維持に必須である。機能障害のあるTORシグナル伝達は、種々のヒト疾患(癌など)に関連している(Metcalf, S.M., Canman, C.E., 他(1997) *Oncogene* 15:1635-1642; Emami, S., Le Flock, N., 他(2001) *FASEB J.* 15:351-361)。また、自己免疫にも関連する(Damoiseaux, J.G., Beijlvel, L.J., 他(1996) *Transplantation* 62:994-1001)。

20

30

【0096】

数種のシクロフィリンイソ酵素が同定されている。例えばシクロフィリンB、シクロフィリンC、ミトコンドリアのマトリクスシクロフィリン、細菌のサイトゾルおよびペリプラズムのPPIアーゼ、並びに、ナチュラルキラー細胞シクロフィリン関連タンパク(シクロフィリン型PPIアーゼドメインを持ち、推定上の腫瘍認識複合体であり、ナチュラルキラー(NK)細胞の機能に関与)を含む。これら細胞は、生得的な細胞免疫応答に、ウイルス感染細胞や癌化細胞を溶解して関与する。NK細胞の特異的標的細胞は、主要組織適合複合体(MHC)クラスI遺伝子の発現を喪失(腫瘍化時に一般的に発生)した細胞であり、この標的化によって、細胞に腫瘍成長減弱の可能性を与える。150-kDaの分子の同定が、ヒトNK細胞の表面で成されている。この分子が持つ或るドメインは、シクロフィリン/ペプチジルプロリンシス-トランス異性化酵素に高度に相同である。このシクロフィリン型タンパクは、推定上の腫瘍認識複合体であるNK腫瘍認識配列(NK-TR)の成分でありうる(Anderson, S.K., Gallinger, S., 他(1993) *Proc. Natl. Acad. Sci. USA* 90:542-546)。NKTR腫瘍認識配

40

50

列は、腫瘍細胞と、大顆粒リンパ球 (LGL) との間の認識を媒介する。LGLは白血球の亜集団であり、活性化した細胞障害性T細胞とナチュラルキラー細胞とからなり、腫瘍標的を破壊できる。NKTR遺伝子のタンパク産物はLGLの表面に見られ、腫瘍標的への結合を促進する。最近、マウスNktr遺伝子とプロモーター領域とが、9番染色体に位置づけられた。この遺伝子は、NK細胞特異的150-kDaタンパク (NK-TR) をコードする。これは、シクロフィリン他の腫瘍応答タンパクに相同である (Simons-Evelyn, M., Young, H.A. および Anderson, S.K. (1997) *Genomics* 40: 94-100)。

【0097】

腫瘍化組織と相互作用する他のタンパクとしては、サイトカイン、例えば腫瘍壊死因子 (TNF) を含む。TNFファミリーのサイトカインの産生はリンパ球とマクロファージとが行い、TNFは、癌化した (腫瘍) 内皮細胞の溶解を生じうる。内皮タンパク1 (Edp1) の同定が、TNF α によって内皮細胞で転写的に活性化されるヒト遺伝子として成されている。また、TNF α 誘導可能Edp1遺伝子が、マウスで同定されている (Swift, S., Blackburn, C., 他 (1998) *Biochim. Biophys. Acta* 1442: 394-398)。

【0098】

発癌遺伝子

多くの発癌遺伝子が同定され特徴付けられている。発癌遺伝子としては、成長因子 (sis等)、erbA、erbB、neu、ros等の受容体、src、yes、fps、abl、met等の細胞内レセプター、mos、raf等のタンパク質セリン/トレオニンキナーゼ、jun、fos、myc、N-myc、myb、ski、relなどの核転写因子、RB、p53等の細胞周期制御タンパク質、mdm2、Cip1、p16、サイクリンD、ras、set、can、sec、gag R10等の突然変異腫瘍抑制遺伝子がある。特に、fosがコードするFOSは、ロイシンジッパー含有リン蛋白であり、細胞の核内にある。FOSは、非共有結合複合体を数種の他のタンパクと形成し、成長促進タンパクの転写を活性化する (Bohmann, D. 他 (1987) *Science* 238:1386-1392; Cohen, D.R. および Curran, T. (1988) *Mol. Cell. Biol.* 8:2063-2069; 並びに van Straaten, F. 他 (1983) *Proc. Natl. Acad. Sci.* 80: 3188-3187)。canは推定上のヒト発癌遺伝子であり、骨髄性白血病誘発に関わる。また、発癌遺伝子としての活性化が、その3'側半分の、他の遺伝子 (setなど) との融合によって成される (von Lindern, M. 他 (1992) *Mol. Cell. Biol.* 12: 3346-3355)。setがコードするSETはホスファターゼ2A (セリン/トレオニンホスファターゼであり、多様な細胞プロセスを調節) の強力な阻害因子であることが示されている (Li, M. 他 (1996) *J. Biol. Chem.* 271:11059-11062)。SETのアフリカツメガエル相同体であるNAP1は、B型サイクリンと特異的に相互作用することがわかり、また、細胞周期調節で必須の役割を果たす (Kellogg, D. R. 他 (1995) *J. Cell Biol.* 130: 661-673)。SECはsecの遺伝子産物であり、この腫瘍性タンパクは、分泌性上皮の腫瘍で活性である (Lane, M.A. 他 (1990) *Nuc. Acids Res.* 18: 3068)。gag R10はロイシンジッパー含有細胞質タンパクで23 kDaであり、ニワトリ胚視神経網膜細胞から同定されており、キメラmRNAであるRAV-1によってコードされている。これは、細胞を継続的細胞増殖に誘導できる (Proux, V. 他 (1996) *J. Biol. Chem.* 271:30790-30797)。S-100ファミリーは、小さな二量体の酸性のカルシウム結合および亜鉛結合タンパクであり、脳で豊富に発現する。これらのタンパク質は、細胞成長および分化、細胞周期調節、並びに代謝制御において重要な役割を果たす (Moncrief, N. D. 他 (1990) *J. Mol. Evol.* 30:522-562; 並びに Wicki, R. 他 (1996) *Biochem. Biophys. Res. Commun.* 227: 594-599)。rad1は酵母タンパクであり、DNA修復およびDNA組換えに関与する (Sunnerhagen, P. 他 (1990) *Mol. Cell. Biol.* 10: 3750-3760)。 α Lフコシダーゼはリソソーム酵素であり、糖タンパクの、フコースと、糖質部分のNアセチルグルコサミンとの間の α 1,6結合を加水分解する。 α Lフコシダーゼの欠損の結果、フコシドーシス (fucosidosis、リソソーム蓄積症) が起きる (Herissat, B. (1991) *Biochem. J.* 280: 309-316)。

【0099】

腫瘍性タンパク質をコードする遺伝子は発癌遺伝子と呼ばれ、これらは、細胞成長および発達を通常は制御する遺伝子に由来する。多くの発癌遺伝子が同定され特徴付けられてい

10

20

30

40

50

る。発癌遺伝子としては、成長因子(sis等)、erbA、erbB、neu、ros等の受容体、src、yes、fps、abl、met等の細胞内レセプター、mos、raf等のタンパク質セリン/トレオニンキナーゼ、jun、fos、myc、N-myc、myb、ski、relなどの核転写因子、RB、p53等の細胞周期制御タンパク質、mdm2、Cipl、p16、サイクリンD、ras、set、can、sec、gag R10等の突然変異腫瘍抑制遺伝子がある。

【0100】

ウイルス発癌遺伝子は、ヒト細胞が数種のウイルスに感染するとヒトゲノムに組込まれる。ウイルス発癌遺伝子の例としては、v-src、v-abl、およびv-fpsを含む。正常遺伝子の、発癌遺伝子への転換はまた、染色体転座によっても起こりうる。フィラデルフィア染色体は、慢性骨髄性白血病と、或るサブセットの急性リンパ芽球性白血病との特徴であり、これは、9番染色体と22番染色体との間の相互転座の結果である。この転座は、プロトオンコジーンc-ablの或る切り詰められた部分を、22番染色体のブレイクポイント（切断点）クラスター領域(bcr)へ移動させる。ハイブリッドc-abl-bcr遺伝子がコードするキメラ蛋白は、チロシンキナーゼ活性を持つ。慢性骨髄性白血病では、このキメラ蛋白が持つ分子量は210 kdであり、急性白血病では、より活性な180 kdのチロシンキナーゼが形成される(Robbins, S.L. 他(1994) Pathologic Basis of Disease, W.B. Saunders Co., Philadelphia PA)。

10

【0101】

Wnt遺伝子ファミリーの分泌シグナル分子は、真核細胞全体で高度に保存されている。Wntファミリーのメンバーは、軟骨テンプレート(cartilage template)内の軟骨細胞分化の調節に関わる。Wnt-5a、Wnt-5b、およびWnt-4遺伝子の発現は、ニワトリ四肢の軟骨形成領域で成される。Wnt-5aの発現は、軟骨膜(早期軟骨テンプレートのすぐ周辺の間葉細胞)で成される。Wnt-5a誤発現は、ニワトリ四肢で、軟骨細胞の成熟と、bone collar形成の開始を遅らせる(Hartmann, C. および Tabin, C.J. (2000) *Development* 127:3141-3159)。

20

【0102】

RRP22タンパク/RAS関連タンパク

シグナル伝達は、それによって細胞が細胞外シグナルに応答する、一般的プロセスである。典型的シグナル伝達経路では、ホルモン、神経伝達物質または成長因子など、シグナル分子の、細胞膜受容体への結合が、細胞内セカンドメッセンジャーの作用と共役される。Gタンパク共役レセプター(GPCR)は、細胞内プロセスの制御を、グアニンヌクレオチド結合タンパク(Gタンパク)を活性化して行う。Gタンパクはヘテロ三量体であり、 α 、 β 、および γ サブユニットからなる。 α サブユニットはグアニンヌクレオチド結合ドメインを含み、GTPアーゼ活性を持つ。GTPが α サブユニットに結合すると、これは β および γ サブユニットから解離し、細胞性標的分子と相互作用する。GTPの、GDPへの加水分解は、このサブユニットの、他のタンパクとの相互作用を制御する分子スイッチとして働く。GDP結合型の α サブユニットは、その細胞標的から解離し、 β および γ サブユニットと再会合する。多数のアクセサリータンパクが、Gタンパク機能の調整を、それらのヌクレオチドリノ酸化状態または膜結合を制御して行う。これら調節分子としては、交換因子(GEF、GDP-GTP交換を刺激)、GTPアーゼ活性化タンパク(GAP、GTP加水分解を促進)、および、グアニンヌクレオチド解離インヒビター(GDI、グアニンヌクレオチド解離を阻害し、GDP結合型を安定化)を含む。Gタンパクを少なくとも5つのサブファミリーに分類しうる。Ras、Rho、Ran、Rab、およびADPリボシル化因子であり、これらは、種々の細胞機能(例えば細胞成長および分化、細胞骨格構成、並びに細胞内小胞輸送および分泌)を調節する。

30

40

【0103】

Rasスーパーファミリーの小GTPアーゼは、広範な細胞シグナル経路の調節に関わる。Rasファミリー蛋白は膜結合タンパクであり、分子スイッチとして作用し、GTPとGDPとに結合し、GTPをGDPへ加水分解する。活性なGTP結合状態では、Rasファミリー蛋白は種々の細胞標的と相互作用し、下流シグナル経路を活性化する。例えばRasサブファミリーのメンバーは、受容体チロシンキナーゼ(RTK)から、一連のセリン/トレオニンキナーゼへのシグナル伝達に必須である。セリン/トレオニンキナーゼは細胞成長および分化を制御する。活

50

性化したRas遺伝子は初めにヒトの癌で見出され、続いての研究は、細胞が成長を続けるか、最終分化に至るかの決定にRas機能が死活的であることを確認した (Barbacid, M. (1987) *Annu. Rev. Biochem.* 56:779-827, Treisman, R. (1994) *Curr. Opin. Genet. Dev.* 4:96-98)。突然変異Rasタンパク (GTPに結合するが加水分解しないRas) は恒久的に活性化されており、継続的細胞増殖または癌を起こす。

【0104】

Rasサブファミリーは、チロシンキナーゼ受容体、非チロシンキナーゼ受容体、および、ヘテロ三量体GPCRからのシグナルを伝達する (Fantl, W. J. 他 (1993) *Annu. Rev. Biochem.* 62:453-481; Woodrow, M. A. 他 (1993) *J. Immunol.* 150:3853-3861; 並びに van Corven, E. J. 他 (1993) *Proc. Natl. Acad. Sci.* 90:1257-1261)。細胞表面レセプターを刺激するとRasが活性化し、Rasが次に、細胞成長および分化を制御する細胞質キナーゼを活性化する。同定された最初のRas標的は、Rafキナーゼであった (Avruch, J. 他 (1994) *Trends Biochem. Sci.* 19:279-283)。RasとRafとの相互作用はセリン/トレオニンキナーゼ群のMAPキナーゼカスケードの活性化を生じ、セリン/トレオニンキナーゼ群は遺伝子発現とタンパク質合成とを制御する重要な転写因子を活性化させる (Barbacid, M. (1987) *Annu. Rev. Biochem.* 56:779-827, Treisman, R. (1994) *Curr. Opin. Genet. Dev.* 4:96-101)。突然変異Rasタンパク (GTPに結合するが加水分解しないRas) は恒久的に活性化されており、継続的細胞増殖および癌を起こす (Bos, J. L. (1989) *Cancer Res.* 49:4682-4689; Grunicke, H. H. and Maly, K. (1993) *Crit. Rev. Oncog.* 4:389-402)。

【0105】

多くの発癌遺伝子が同定され特徴付けられている。発癌遺伝子としては、成長因子 (sis等)、erbA、erbB、neu、ros等の受容体、src、yes、fps、abl、met等の細胞内レセプター、mos、raf等のタンパク質セリン/トレオニンキナーゼ、jun、fos、myc、N-myc、myb、ski、relなどの核転写因子、RB、p53等の細胞周期制御タンパク質、mdm2、Cip1、p16、サイクリンD、ras、set、can、sec、gag R10等の突然変異腫瘍抑制遺伝子がある。特に、fosがコードするFOSは、ロイシンジッパー含有リン蛋白であり、細胞の核内にある。FOSは、非共有結合複合体を数種の他のタンパクと形成し、成長促進タンパクの転写を活性化する (Bohmann, D. 他 (1987) *Science* 238:1386-1392; Cohen, D.R. および Curran, T. (1988) *Mol. Cell. Biol.* 8:2063-2069; 並びに van Straaten, F. 他 (1983) *Proc. Natl. Acad. Sci.* 80: 3188-3187)。canは推定上のヒト発癌遺伝子であり、骨髄性白血病誘発に関わる。また、発癌遺伝子としての活性化が、その3'側半分の、他の遺伝子 (setなど) との融合によって成される (von Lindern, M. 他 (1992) *Mol. Cell. Biol.* 12: 3346-3355)。setがコードするSETはホスファターゼ2A (セリン/トレオニンホスファターゼであり、多様な細胞プロセスを調節) の強力な阻害因子であることが示されている (Li, M. 他 (1996) *J. Biol. Chem.* 271:11059-11062)。SETのアフリカツメガエル相同体であるNAP1は、B型サイクリンと特異的に相互作用することがわかり、また、細胞周期調節で必須の役割を果たす (Kellogg, D. R. 他 (1995) *J. Cell Biol.* 130: 661-673)。SECはsecの遺伝子産物であり、この腫瘍性タンパクは、分泌性上皮の腫瘍で活性である (Lane, M.A. 他 (1990) *Nuc. Acids Res.* 18: 3068)。gag R10はロイシンジッパー含有細胞質タンパクで23 kDaであり、ニワトリ胚視神経網膜細胞から同定されており、キメラmRNAであるRAV-1によってコードされている。これは、細胞を継続的細胞増殖に誘導できる (Proux, V. 他 (1996) *J. Biol. Chem.* 271:30790-30797)。S-100ファミリーは、小さな二量体の酸性のカルシウム結合および亜鉛結合タンパクであり、脳で豊富に発現する。これらのタンパク質は、細胞成長および分化、細胞周期調節、並びに代謝制御において重要な役割を果たす (Moncrief, N. D. 他 (1990) *J. Mol. Evol.* 30:522-562; 並びに Wicki, R. 他 (1996) *Biochem. Biophys. Res. Commun.* 227: 594-599)。rad1は酵母タンパクであり、DNA修復およびDNA組換えに関与する (Sunnerhagen, P. 他 (1990) *Mol. Cell. Biol.* 10: 3750-3760)。αLフコシダーゼはリソソーム酵素であり、糖タンパクの、フコースと、糖質部分のNアセチルグルコサミンとの間のα1,6結合を加水分解する。αLフコシダーゼの欠損の結果、フコシドーシス (fucosidosis、リソソーム蓄積症) が起きる (Herissat, B. (1991) *Biochem. J.* 280: 3

10

20

30

40

50

09-316)。

【0106】

Rasは、他のシグナル経路の調節を、種々の細胞標的と直接に相互作用して行う(Katz, M. E. および McCormick, F. (1997) Curr. Opin. Genet. Dev. 7:75-79)。このような標的の1つであるRalGDSは、Ras様GTPアーゼであるRalのグアニンヌクレオチド解離刺激因子である(Albright, C. F. 他(1993) EMBO J. 12:339-347)。RalGDSは、RasとRalとのシグナル経路を共役する。上皮成長因子(EGF)は、RalGDSとRasとの会合を、哺乳類細胞で刺激する。この会合は、RalGDSのGEF作用を活性化する(Kikuchi, A. および Williams, L. T. (1996) J. Biol. Chem. 271:588-594; Urano, T. 他(1996) EMBO J. 15:810-816)。RalGDSによるRal活性化は、Srcの活性化を生じる。Srcはチロシンキナーゼであり、他の分子(例えば転写因子および、アクチン細胞骨格の諸成分)をリン酸化する(Goi, T. 他(2000) EMBO J. 19:623-630)。Ralは、多数のシグナル分子と相互作用する。例えばRal結合タンパク(Rho様GTPアーゼのGAP)、Cdc42およびRac(細胞骨格再編成を調節)、並びに、ホスホリパーゼD1(小胞輸送に関与)を含む(Feig, L. A. 他(1996) Trends Biochem. Sci. 21:438-441; Voss, M. 他 (1999) J. Biol. Chem. 274:34691-34698)。

10

【0107】

Norelの同定は、酵母2ハイブリッドスクリーニングから、RasおよびRas関連タンパクであるRap1bと相互作用するタンパクとして成された(Vavvas, D. 他(1998) J. Biol. Chem. 273:5439-5442)。Norelは塩基性タンパク(pI=9.4)で413アミノ酸から成り、システイン-ヒスチジンリッチ領域(予測ではジアシルグリセロール/ホルボールエステル結合部位)、N末端のプロリンリッチ領域(SH3結合ドメインの可能性)、および、C末端のRas/Rap結合ドメインを持つ。Norelは、*in vitro*で、GTP依存的にRasと結合する。*in vivo*試験が示すところでは、NorelのRasとの会合は、COS-7細胞でEGFおよび12-O-テトラデカノイルホルボール-13アセテート活性化に、またはKB細胞でEGFに依存する。

20

【0108】

Ras他のGタンパクは、免疫炎症応答の調節において役割を果たす。顆粒球(好塩基球、好酸球、好中球を含む)は、炎症で決定的役割を果たす。好酸球は毒性顆粒タンパクを放出し、これは微生物を殺す。また、プロスタグランジン、ロイコトリエン、およびサイトカインを分泌し、これらは炎症応答を増幅する。これらはアレルギー反応の炎症を持続させ、その機能不全は、喘息などアレルギー疾患を生じうる。インターロイキン5はサイトカインであり、好酸球の成長、活性化、および生存を調節する。IL-5の、好酸球でのシグナル伝達機序には、Ras-MAPキナーゼ経路とJak-Stat経路とが関わる(Pazdrak, K. 他(1995) J. Exp. Med. 181:1827-1834; Adachi, T. および Ala, R. (1998) Am. J. Physiol. 275:C623-633)。RasによるRaf-1キナーゼ活性化は、好酸球脱顆粒に関連付けられる。

30

【0109】

好中球は炎症部位に遊走し、ここで病原体の除去を貪食で行い、毒性産物の放出をその顆粒から行って、これが微生物を殺す。GタンパクにはRas、Ral、Rac1、およびRap1を含み、これらは好中球機能を調節する(M'Rabet, L. 他(1999) J. Biol. Chem. 274:21847-21852)。Rac1は、好中球の呼吸バーストに関わりうる。RasとRap1との活性化は、走化因子であるホルミルメチオニンロイシンフェニルアラニン(fMLP)、脂質メディエータである血小板活性化因子(PAF)、および、サイトカインである顆粒球マクロファージコロニー刺激因子(GM-CSF)に応じて成される。RasとRap1との双方は、好中球活性化で役割を果たすようである。Ralの活性化はfMLPとPAFとによって成されるがGM-CSFでは成されず、Ralは走化性、貪食、または脱顆粒に関与しうる。好中球機能の障害には、種々の炎症疾患と自己免疫疾患とが関わる。

40

【0110】

RRP22は、その発現が中枢神経系に限られる新規な亜群を定義する。RRP22遺伝子の所在は、22番染色体のCpGリッチq12領域内の、40-kbの領域であり、EWSおよびBAM22遺伝子が境界を成す(Zucman-Rossi, J. 他(1996) Genomics 38:247-254)。

【0111】

50

Rasファミリータンパクの活性化の触媒はグアニンヌクレオチド交換因子(GEF)が行い、GEFは、結合したGDPの解離と、その後のGTPの結合とを触媒する。最近発見されたRalGEF様タンパクであるRGL3は、Rasと近縁タンパクRitの双方と相互作用する。構成的に活性なRitは、Ras同様、癌原性転換を誘発しうるが、Ritは大抵の既知Rasエフェクター蛋白と相互作用しないので、新規な細胞標的が、Rit転換活性に関わる可能性がある。RGL3はRasとRitの双方と相互作用するので、これら蛋白の下流エフェクターとして作用する可能性がある(Shao, H. および Andres, D.A. (2000) J. Biol. Chem. 275:26914-26924)。

【0112】

腫瘍抗原

腫瘍抗原は細胞表面分子であり、腫瘍細胞では非腫瘍組織と異なる発現をする。腫瘍抗原は、腫瘍細胞を免疫的に正常細胞と異なるものとし、ヒトの癌での潜在的診断材料である。数種のモノクローナル抗体が、癌細胞と特異的に反応することが同定された(例えばT細胞急性リンパ芽球白血病および神経芽腫)(Minegishi 他(1989) Leukemia Res.13:43-51; Takagi他(1995) Int. J. Cancer 61:706-715)。また、HER2遺伝子の、乳腺腫瘍での高レベル発現の発見は、治療処置の開発を導いた(Liu 他(1992) Oncogene 7:1027-1032; Kern (1993) Am. J. Respir. Cell Mol. Biol. 9:448-454)。腫瘍抗原は細胞表面に見られ、特徴は、膜タンパクまたは糖タンパクであることである。例えばMAGE遺伝子がコードするファミリーの腫瘍抗原の認識は、メラノーマ細胞表面で、自家細胞溶解性Tリンパ球が行う。12種のヒトMAGE遺伝子が単離された内、半分は、種々の組織学タイプの腫瘍で差次的に発現する(De Plaen 他(1994) Immunogenetics 40:360-369)。しかし、12種のMAGE遺伝子のどの発現も、健常組織では精巣と胎盤と以外では見られない。

【0113】

乳癌

180,000例を超える新規発症の乳癌が毎年診断されており、乳癌による死亡率は45～54歳の女性の全死亡の10%に近い(K. Gish (1999) AWIS Magazine 28:7-10)。ただし、限局性乳癌の早期診断に基づく生存率はきわめて高い(97%)。これに対して、疾患のステージが進行し腫瘍が乳房外に進展した場合は22%である。現在の臨床乳癌検診の手法は感度と特異性とを欠いており、乳癌の包括的な遺伝子発現プロファイルを開発する努力がなされている。このプロファイルを従来のスクリーニング法とともに用いれば、乳癌の診断と予後診断とを向上し得る(Perou, C.M. 他(2000) Nature 406:747-752)。

【0114】

2種の遺伝子、BRCA1とBRCA2との突然変異が、女性の乳癌の大きな素因であることが知られており、この変異は親から子へ受け渡されるようである(Gish, K. (1999) AWIS Magazine 28:7-10)。しかし、このタイプの遺伝性乳癌はわずかに乳癌の約5%～9%であり、大多数の乳癌の原因は非遺伝性突然変異であり、乳腺上皮細胞での変異である。

【0115】

上皮成長因子(EGF)とその受容体であるEGFRとの発現と、ヒト乳癌との関係は、特に良く研究されている(Khazaie, K. 他(1993) Cancer and Metastasis Rev. 12:255-274とその中の参考文献に、この分野が概説されている)。EGFRの過剰発現、特にエストロゲン受容体の下方調節と結びついた発現は、乳癌患者の不良な予後のマーカーである。また、乳腺腫瘍転移におけるEGFR発現は、しばしば原発腫瘍に比して上昇することから、EGFRは腫瘍の進行と転移とに関わることを示唆している。これを支持する証拠が蓄積されている。それは、EGFが、転移の潜在能に関する細胞機能に対する効果を持つという証拠である。この機能とは例えば、細胞運動性、走化性、分泌、および分化である。erbB受容体ファミリー(EGFRはその1つである)の他のメンバーの発現の変化も、乳癌に関わるとされている。erbB受容体(例えばHER-2/neu, HER-3およびHER-4)およびそれらのリガンドの乳癌における豊富さは乳癌の発生機序におけるそれらの機能的な重要性を示しており、したがって乳癌治療の標的を提供しうる(Bacus, S. S. 他(1994) Am. J. Clin. Pathol. 102:S13-S24)。別の既知の乳癌マーカーとしては、或るヒトの分泌型frizzled protein mRNA(乳腺腫瘍では下方調節される)、マトリクスGlaタンパク質(ヒト乳癌細胞で過剰発現)、Drg1す

なわちRTP(この遺伝子の発現は結腸腫瘍、乳腺腫瘍、前立腺腫瘍で低下)、maspin(この腫瘍抑制遺伝子は浸潤性乳癌で下方調節)、およびCaN19(S100蛋白ファミリーのメンバーであり、このファミリーは全て、乳癌細胞では正常乳腺上皮細胞に比して下方調節される)を含む(Zhou, Z. 他(1998) Int. J. Cancer 78:95-99; Chen, L. 他(1990) Oncogene 5:1391-1395; Ulrix, W. et al(1999) FEBS Lett 455:23-26; Sager, R. 他(1996) Curr. Top. Microbiol. Immunol. 213:51-64;およびLee, S. W. 他(1992) Proc. Natl. Acad. Sci. USA 89:2504-2508)。

【0116】

種々の病期のヒト乳癌の乳腺上皮細胞に由来する細胞株は、悪性転換と腫瘍進行とのプロセスを研究するのに有用なモデルを提供する。その理由は、これらの細胞株が、その親腫瘍の多くの特性を、長い培養期間にわたり保持することが示されているからである(Wistuba, I.I. 他(1998) Clin.Cancer Res. 4:2931-2938)。このようなモデルが特に有用なのは、ヒト乳腺上皮細胞の表現型と分子との特徴を、種々のステージの悪性転換について比較する場合である。

10

【0117】

腫瘍抑制因子

腫瘍抑制遺伝子の一般的定義は、遺伝エレメントであってその喪失または不活化が、細胞増殖の調節解除と、癌の発生機序と進行とに寄与する要素である。腫瘍抑制遺伝子の正常な機能は、細胞成長の制御または阻害をストレスに応じて行い、細胞の増殖寿命を限定することである。数種の腫瘍抑制遺伝子が同定されており、これには網膜芽腫(Rb)タンパク、p53、並びに乳癌1および2タンパク(BRCA1およびBRCA2)をコードする遺伝子を含む。これら遺伝子の突然変異は、数種の癌の発症への、後天性および先天性の遺伝素因と関連する。

20

【0118】

p53の、癌の発生機序における役割は、広範に研究されている(概説はAggarwal, M. L. 他(1998) J. Biol. Chem. 273:1-4; Levine, A. (1997) Cell 88:323-331)。全てのヒト癌の約50%には、p53遺伝子の突然変異を含む。これら突然変異の結果、機能的p53の不在や、より多くは、過剰発現される欠損型のp53を生じる。p53は転写因子であり、DNA結合に必要な中心コアドメインを持つ。p53での大抵の癌関連突然変異は、このドメインに局在する。正常な増殖細胞では、p53は発現が低レベルであり、迅速に分解される。p53の発現と活性との誘発は、DNA損傷、頓座有糸分裂、その他のストレス性の刺激に応じて起こる。これらの場合、p53はアポトーシスを誘発させ又は、細胞成長をストレスが除去されるまで停止させる。p53活性の下流エフェクターとしては、アポトーシス特異的タンパクおよび細胞周期調節タンパク(例えばRb)、発癌遺伝子産物、サイクリン、および細胞周期依存キナーゼなどを含む。

30

【0119】

転移抑制遺伝子であるKAI1(CD82)は、腫瘍抑制遺伝子p53に近縁であるとの報告がある。KAI1は、その発現の低下時に、ヒト前立腺癌の進行、また、おそらく肺癌と乳癌との進行に関わる。KAI1は、構造的に特有な、白血球表面糖タンパク質のファミリーのメンバーをコードする。このファミリーは、テトラスパン膜貫通タンパク質ファミリーまたは膜貫通4スーパーファミリー(TM4SF)として知られる。理由は、このファミリーのメンバーが、形質膜を4回貫通するからである。このファミリーは、内在性膜タンパクからなる。これらはN末端膜固着ドメインを持ち、この機能は、膜アンカーとして、また、タンパク生合成時の移行シグナルとしてのものである。N末端膜固着ドメインは、生合成時に切断されない。TM4SFタンパクは、3つの更なる膜貫通領域群、7個以上の保存システイン残基を持ち、同様のサイズであり(218~284残基)、全てが、大きな細胞外親水性ドメイン(3つの潜在的Nグリコシル化部位を持つ)を有する。プロモーター領域には種々の転写因子の、多くの推定結合モチーフを持ち、5つのAP2部位と、9つのSpI部位を含む。KAI1と7種の他のTM4SFメンバーとの遺伝子構造比較が示すところでは、予測されたタンパクの種々の構造ドメインに関連したスプライス部位が保存されている。これが示唆するところでは、こ

40

50

れら遺伝子は進化的に近縁で、遺伝子重複および分岐進化で生じた(Levy, S. 他(1991) J. Biol. Chem. 266:14597-14602; Dong, J.T. 他 (1995) Science 268:884-886; Dong, J.T. 他 (1997) Genomics 41:25-32)。

【0120】

ロイシンリッチ遺伝子グリオーマ不活化(LGI1)タンパクは、受容体および接着タンパクとして機能する多数の膜貫通および細胞外タンパクとの相同性を共有する。LGI1のコードは或るLRR(ロイシンリッチ、リピート含有)遺伝子が行い、10q24にマップする。LGI1は4つのLRR(システインリッチ領域に隣接)と、1つの膜貫通ドメインとを持つ(Somerville, R.P., 他(2000) Mamm.Genome 11:622-627)。LGI1発現は、主に神経組織、特に脳で見られる。腫瘍抑制作用の喪失は、LGI1タンパクの不活化で見られ、これは、悪性グリオーマの、低から高グレードの腫瘍の移行時に起こる。LGI1発現の、低グレード脳腫瘍での低下と、悪性グリオーマでの発現の顕著な低下または不在とが示唆するところでは、LGI1は、グリア腫瘍進行の診断に用い得る(Chernova, O.B., 他(1998) Oncogene 17: 2873-2881)。

10

【0121】

ST13腫瘍抑制因子の同定は、正常な粘膜組織のcDNAと、結腸直腸癌組織のmRNAとの間のサブトラクションハイブリダイゼーションで、結腸直腸癌に関する因子のスクリーニングで成された(Cao, J. 他(1997) J. Cancer Res. Clin. Oncol. 123:447-451)。ST13は、ヒト結腸直腸癌で下方制御される。

【0122】

von Hippel-Lindau (フォン・ヒッペル・リンドウ、VHL)腫瘍抑制遺伝子の突然変異は、網膜および中枢神経系の血管芽腫、明細胞腎癌、および褐色細胞腫に関わる(Hoffman, M. 他(2001) Hum. Mol. Genet. 10:1019-1027; Kamada, M. (2001) Cancer Res. 61:4184-4189)。腫瘍進行は、VHL遺伝子の欠損または不活化に関連する。VHLは、トランスフォーミング成長因子 α 、GLUT-1グルコース輸送体、および血管内皮成長因子の発現を調節する。VHLタンパクは、elongin B、elongin C、Cul2およびRbx1と会合して複合体を形成し、転写アクチベーターである低酸素誘導因子(HIF)を調節する。HIFは血管新生に関わる遺伝子(例えば血管内皮成長因子と血小板由来成長因子B)を誘導する。HIFの制御の喪失はVHLの欠損に起因し、その結果、血管新生ペプチドの過剰な産生を生じる。VHLは、血管新生の阻害、細胞周期制御、フィブロネクチンマトリックスアセンブリ、細胞接着、並びにタンパク分解において役割を果たしうる。

20

30

【0123】

腫瘍抑制遺伝子の突然変異は多くの癌の一般的特徴であり、しばしば、腫瘍の発生機序と進行との決定的ステップに影響するようである。従ってChang, F. 他(1995; J. Clin. Oncol. 13:1009-1022)が示唆するところでは、この遺伝子または、発現するタンパクの抗体を用いて、1) 癌のリスクの高い患者のスクリーニング、2) 従来の方法で成される診断の補助、および、3) 個々のがん患者の予後の評価が可能である。また、Hamada, K他(1996; Cancer Res.56:3047-3054)は、p53の、子宮頸癌細胞への、アデノウイルスベクターでの導入を、子宮頸癌の実験的治療として試験中である。

【0124】

PRドメイン遺伝子は最近、ヒト腫瘍化で役割を果たすことが認識された。PRドメイン遺伝子は通常、2種のタンパク産物を産生する。PRプラス産物(PRドメインを持つ)と、PRマイナス産物(PRドメインを欠く)である。癌細胞では、PRプラスは破壊または過剰発現され、PRマイナスは存在または過剰発現する。これら2種のタンパクの量の不均衡は、悪性疾患の重要な一因の様である(Jiang, G.L. および Huang, S. (2000) Histol. Histopathol. 15:109-117)。

40

【0125】

ヒトにおける多くの腫瘍疾患は、不適切な遺伝子転写によって起こり得る。悪性の細胞成長は、腫瘍促進遺伝子の過剰な発現又は、腫瘍抑制遺伝子の不十分な発現によって起こり得る(Cleary, M. L. (1992) Cancer Surv. 15:89-104)。染色体転座によっても、或る遺伝子のコード配列を別の無関係遺伝子の調節領域と融合させる、キメラ座(chimeric l

50

oci) が生成され得る。或る重要なクラスの転写調節因子が、Znフィンガー蛋白である。Znフィンガーモチーフは亜鉛イオンに結合する。このモチーフは一般に、周期的に配置されたシステイン残基とヒスチジン残基からなる約30アミノ酸のタンデムリピート群を持つ。この配列パターンの例としては、C2H2タイプ、C4タイプおよびC3HC4タイプZnフィンガー群、並びにPHDドメインが挙げられる (Lewin, 前出; Aasland, R., 他(1995) Trends Biochem. Sci. 20:56-59)。臨床的に関連するZnフィンガー蛋白質の一つがWT1であるが、これはウィルムス腫瘍を持つ子供において不活性化されている腫瘍抑制タンパク質である。大細胞リンパ腫で重要な役割を果たすオンコジーンbcl-6もZnフィンガー蛋白質である (Papavassiliou, A.G. (1995) N. Engl. J. Med. 332:45-47)。

【0126】

10

腫瘍応答タンパク質

癌(別称は新形成)の特徴は、継続的で無制御な細胞増殖である。癌は、3つのカテゴリーに分類できる。癌腫、肉腫、白血病である。癌腫は柔らかい上皮細胞の悪性成長であり、周辺組織を浸潤し、転移性腫瘍を生じうる。肉腫は、上皮起源のものや、結合組織に起因するものがある。白血病は造血組織の進行性悪性疾患であり、特徴は、白血球とその前駆体との増殖であり、骨髄性(顆粒球または単球に由来)或いはリンパ球性(リンパ球に由来)に分類されうる。腫瘍化とは、腫瘍の発端からの成長の進行を指す。悪性細胞は、起源の組織内の正常な細胞に極めて類似のものや、未分化(低分化)なものがある。腫瘍細胞は、少数の核や、1個の大きな多形核を持ちうる。未分化細胞は、無秩序な塊に成長しうる。塊は血管化が乏しく、その結果、大きな領域が虚血壊死となっている。分化した新生細胞は、起源の組織と同じタンパク質を分泌しうる。癌は成長し、周辺組織に浸潤し、侵襲し、破壊する。これは、体腔または体表への直接播種によって、リンパ性転移によって、または血行性転移によって成される。癌は、主要な公衆保健の関心事であり続けており、現行の予防策と治療とは、大半の患者の要求を満たしていない。腫瘍化の新形成プロセスの理解は、予後のおよび診断的重要性を持つ分子マーカー群の同定によって補助されうる。

20

【0127】

癌治療の現行の形態としては、免疫抑制剤の使用を含む(Morisaki, T. 他(2000) Anticancer Res. 20: 3363-3373; Georger, B. 他(2001) Cancer Res. 61:1527-1532)。細胞シグナリングに関わるタンパク質の同定は、また特に、免疫抑制剤の受容体として作用するタンパク質の同定は、抗腫瘍剤の開発を促進しうる。例えばイムノフィリンは、原核生物と真核生物との双方に見られる保存されたタンパク質の1ファミリーであり、免疫抑制剤に、種々の程度の特異性で結合する。このようなimmunophilicタンパク質の1グループが、ペプチジルプロリンシスートランス異性化酵素(EC 5.2.1.8)ファミリー (PPIアーゼ、rotamase) である。これら酵素は、初めにブタ腎臓皮質から単離され、これらはタンパク折り畳みの加速を、オリゴペプチド内のプロリンイミドペプチド結合のシス-トランス異性化を触媒して行う(Fischer, G. および Schmid, F.X. (1990) Biochemistry 29: 2205-2212)。イムノフィリンファミリーには、シクロフィリン (例えばペプチジルプロリン異性化酵素A、即ちPPIA) とFK結合タンパク(即ちFKBP)とのサブファミリーを含む。シクロフィリンは多機能レセプター蛋白であり、これは、シグナル伝達活性(例えばシクロスポリンが媒介する伝達)に関わる。各ファミリーのPPIアーゼドメインは、種の間で高度に保存される。構造的には異なるが、これら多機能レセプター蛋白は、多数のシグナル伝達系路に関わる。また、折り畳みイベントと輸送イベントとに関連付けられている。

30

40

【0128】

イムノフィリン蛋白であるシクロフィリンは、免疫抑制剤であるシクロスポリンAと結合する。FKBPは別のイムノフィリンであり、FK506(またはラパマイシン)と結合する。ラパマイシンは免疫抑制剤であり、細胞を成長のG₁期で停止させ、アポトーシスを誘発する。シクロフィリン同様、このマクロライド抗生物質(産生は放線菌 (*Streptomyces tsukubae nsis*) に由来)の作用は、遍在性で主にサイトゾル性のイムノフィリン受容体に結合して成される。これらイムノフィリン/免疫抑制剤複合体(例えばシクロフィリンA/シクロスポリンA(CypA/CsA)、およびFKBP12/FK506)は、その治療成績の達成を、ホスファターゼで

50

あるカルシニューリン(カルシウム／カルモジュリン依存性タンパク質キナーゼであり、T細胞活性化に関与)を阻害して行う(Hamilton, G.S. および Steiner, J.P. (1998) J. Med. Chem. 41: 5119-5143)。ネズミ fkbp51 遺伝子の豊富な発現は、免疫組織(例えば胸腺およびTリンパ球)に見られる(Baughman, G. 他(1995) Molec. Cell. Biol. 15: 4395-4402)。FKBP12/ラパマイシン指示型免疫抑制は、TOR(酵母)またはFRAP(FKBP12-ラパマイシン会合タンパク、哺乳類細胞内)に結合して起きる。これらはラパマイシンのキナーゼ標的であり、正常な細胞成長パターンの維持に必須である。機能障害性のあるTORシグナル伝達は、種々のヒト疾患(癌など)に関連している(Metcalfe, S.M. 他(1997) Oncogene 15:1635-1642; Emami, S. 他(2001) FASEB J. 15:351-361)。また、自己免疫にも関連する(Damoiseaux, J.G. 他(1996) Transplantation 62:994-1001)。

10

【0129】

数種のシクロフィリンイソ酵素が同定されている。例えばシクロフィリンB、シクロフィリンC、ミトコンドリアのマトリクスシクロフィリン、細菌のサイトゾルおよびペリプラズムPPIアーゼ、並びに、ナチュラルキラー細胞シクロフィリン関連タンパク(シクロフィリン型PPIアーゼドメインを持ち、推定上の腫瘍認識複合体であり、ナチュラルキラー(NK)細胞の機能に関与)を含む。これら細胞は、生得的な細胞免疫応答に、ウイルス感染細胞や癌化細胞を溶解して関与する。NK細胞の特異的標的細胞は、主要組織適合複合体(MHC)クラスI遺伝子の発現を喪失(腫瘍化時に一般的に発生)した細胞であり、この標的化によって、細胞に腫瘍成長減弱の可能性を与える。150-kDaの分子の同定が、ヒトNK細胞の表面で成されている。この分子が持つ或るドメインは、シクロフィリン/ペプチジルプロリルシス-トランス異性化酵素に高度に相同である。このシクロフィリン型タンパクは、推定上の腫瘍認識複合体であるNK腫瘍認識配列(NK-TR)の成分でありうる(Anderson, S.K. 他(1993) Proc. Natl. Acad. Sci. USA 90:542-546)。NKTR腫瘍認識配列は、腫瘍細胞と、大顆粒リンパ球(LGL)との間の認識を媒介する。LGLは白血球の亜集団であり、活性化した細胞障害性T細胞とナチュラルキラー細胞とからなり、腫瘍標的を破壊できる。NKTR遺伝子のタンパク産物はLGLの表面に見られ、腫瘍標的への結合を促進する。最近、マウスNktr遺伝子とプロモーター領域とが、9番染色体に位置づけられた。この遺伝子は、NK細胞特異的150-kDaタンパク(NK-TR)をコードする。これは、シクロフィリン他の腫瘍応答タンパクに相同である(Simons-Evelyn, M.他(1997) Genomics 40:94-100)。

20

【0130】

腫瘍化組織と相互作用する他のタンパクとしては、サイトカイン、例えば腫瘍壊死因子(TNF)を含む。TNFファミリーのサイトカインの産生はリンパ球とマクロファージとが行い、TNFは、癌化した(腫瘍)内皮細胞の溶解を生じうる。内皮タンパク1(Edp1)の同定が、TNFαによって内皮細胞で転写的に活性化されるヒト遺伝子として成されている。また、TNFα誘導可能Edp1遺伝子が、マウスで同定されている(Swift, S. 他(1998) Biochim. Biophys. Acta 1442: 394-398)。

30

【0131】

加齢および老化

加齢プロセスまたは老化の研究は、多数の特徴的な細胞性変化と分子変化とを示している(Fauci 他(1998) Harrison's Principles of Internal Medicine, McGraw-Hill, New York NY, 37ページ)。これら特徴としては、増加した染色体構造異常、DNA架橋、DNAの一本鎖切断の発生、DNAメチル化の喪失、およびテロメア領域の分解を含む。これらDNA変化に加え、蛋白の翻訳後変容が増し、これには脱アミド化、酸化、架橋、および非酵素的糖付加を含む。尚一層の分子変化の発生が、加齢細胞のミトコンドリアに、構造の劣化を通して起こる。これらの変化は最終的に、身体の各器官の機能低下に寄与する。

40

【0132】

肺癌

肺癌は、米国男性の癌死の主因であり、女性の癌死の第2の原因である。肺癌は、4つの組織病理的に異なる群に分けられる。3群(扁平上皮癌、腺癌、大細胞癌)は、非小細胞肺癌(NSCLC)に分類される。第4群の癌は、小細胞肺癌(SCLC)という。3番染色体での欠失は

50

この疾患に一般的であり、この領域に、腫瘍抑制遺伝子の存在を示すと思われる。K-rasの活性化突然変異は肺癌で一般的に見られ、この疾患のマウスモデルの1つの基礎である。

【0133】

ステロイドホルモン

グルココルチコイドは天然ホルモンであり、薬理量で投与すると、炎症および免疫応答を防止または抑制する。分子レベルでは、非結合グルココルチコイドは容易に細胞膜を越え、特異的細胞質レセプターに高親和性の結合をする。結合後、転写に影響し、最終的に、タンパク合成に影響する。この結果には、炎症部位の白血球浸潤阻害、炎症応答メディエータ機能妨害、および液性免疫応答抑制を含みうる。コルチコステロイドの抗炎症性作用は、ホスホリパーゼA2阻害タンパク類、総称リポコルチンに関わると見られる。リポコルチンは次に、炎症の強力なメディエータ(たとえばプロスタグランジンやロイコトリエン)の生合成を、前駆体分子アラキドン酸の放出を阻害して制御する。更に、コルチコステロイドは、好酸球、好塩基球、および気道上皮細胞機能の阻害を、炎症応答を媒介するサイトカインを調節して行う。これらは、炎症部位の白血球浸潤阻害、炎症応答メディエータ機能妨害、および液性免疫応答抑制を行う。コルチコステロイドを用いて、アレルギー、喘息、関節炎、および皮膚症状が治療される。ベクロメタゾンは合成グルココルチコイドであり、これを用いて、ステロイド依存喘息を治療し、アレルギー性または非アレルギー性(血管運動)鼻炎に伴う症状を軽減し、外科切除後の再発鼻ポリープを予防している。鼻腔内ベクロメタゾンの抗炎症性効果および血管収縮効果は、ヒドロコルチゾンの効果より5000倍大きい。

【0134】

発現プロファイル作成

アレイ技術は、単一の多型遺伝子の発現や、多数の関連遺伝子または無関係の遺伝子の発現プロファイルを探求する、簡単な方法を提供し得る。単一遺伝子の発現を試験するときは、アレイを用いて、或る特定遺伝子又はその変異体の発現を検出する。発現プロファイルを試験するときは、アレイは次のような遺伝子を同定するプラットフォームを提供する。即ちどの遺伝子が組織特異的か、毒性アッセイにおいてテストされる物質に影響されるか、シグナル伝達カスケードの一部であるか、ハウスキーピング機能を実行するか、又は、特定の遺伝的素因や、条件、疾患、又は障害に、特異的に関連する遺伝子であるかの同定である。

【0135】

新規の細胞成長、分化および細胞死関連タンパク質、およびそれらをコードするポリヌクレオチドの発見により、新規の組成物を提供することで当分野の要望に応えることができる。この新規の組成物は、癌など細胞増殖異常、発達障害、神経疾患、自己免疫/炎症の疾患、生殖障害、および胎盤障害の、診断・治療・予防において有用であり、また、細胞成長、分化および細胞死関連タンパク質の核酸配列及びアミノ酸配列の発現における外因性化合物の影響についての評価にも有用である。

【発明の開示】

【発明の効果】

【0136】

本発明は、総称して「CGDD」、個別にはそれぞれ「CGDD-1」、「CGDD-2」、「CGDD-3」、「CGDD-4」、「CGDD-5」、「CGDD-6」、「CGDD-7」、「CGDD-8」、「CGDD-9」、「CGDD-10」、「CGDD-11」、「CGDD-12」、「CGDD-13」、「CGDD-14」、「CGDD-15」、「CGDD-16」、「CGDD-17」、「CGDD-18」、「CGDD-19」、「CGDD-20」、「CGDD-21」と呼ぶ細胞成長、分化および細胞死関連タンパク質である精製されたポリペプチドを提供する。或る態様において本発明は、(a) SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチド、(b) SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列との少なくとも90%の同一性を持つ天然アミノ酸配列を有するポリペプチド、(c) SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチドの生物学的

活性断片、および（d）SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチドの免疫原性断片、からなる群から選択した、単離されたポリペプチドを提供する。一実施例で本発明は、SEQ ID NO:1-21のアミノ酸配列を持つ、単離されたポリペプチドを提供する。

【0137】

また、本発明は（a）SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチド、（b）SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列との少なくとも90%の同一性を持つ或る天然アミノ酸配列を有するポリペプチド、（c）SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチドの生物学的活性断片、または（d）SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリヌクレオチドの免疫原性断片、からなる群から選択した或るポリペプチドをコードする、単離されたポリヌクレオチドを提供する。一実施態様では、該ポリヌクレオチドは、SEQ ID NO:1-21からなる群から選択した或るポリペプチドをコードする。別の実施態様では、ポリヌクレオチドはSEQ ID NO:22-42からなる群から選択される。

【0138】

本発明は更に、（a）SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチド、（b）SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列との少なくとも90%の同一性を有する或る天然アミノ酸配列を持つポリペプチド、（c）SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を有するポリペプチドの生物学的活性断片、および（d）SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチドの免疫原性断片、からなる群から選択した或るポリペプチドをコードするポリヌクレオチドと機能的に連結したプロモーター配列を持つ組換えポリヌクレオチドを提供する。一実施態様で本発明は、この組換えポリヌクレオチドを用いて形質転換した細胞を提供する。別の実施態様で本発明は、この組換えポリヌクレオチドを持つ遺伝形質転換体を提供する。

【0139】

更に本発明は、（a）SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチドと、（b）SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列との少なくとも90%以上の同一性を有する或る天然アミノ酸配列を持つポリペプチドと、（c）SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチドの生物学的活性断片と、（d）SEQ ID NO:1-21からなる群から選択したアミノ酸配列を持つポリペプチドの免疫原性断片、とからなる群から選択したポリペプチドを製造する一方法を提供する。製造方法は、（a）或る細胞を該ポリペプチドの発現に適した条件下で培養し、この細胞を組換えポリヌクレオチドを用いて形質転換する過程と、（b）そのように発現したポリペプチドを回収する過程からなる。この組換えポリヌクレオチドは、該ポリペプチドをコードするポリヌクレオチドに対し機能的に連結したプロモーター配列を持つ。

【0140】

本発明は更に、（a）SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチド、（b）SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列との少なくとも90%の同一性を有する或る天然アミノ酸配列を持つポリペプチド、（c）SEQ ID NO:1-21からなる群から選択したアミノ酸配列を持つポリペプチドの生物学的活性断片、および（d）SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチドの免疫原性断片、からなる群から選択した或るポリペプチドに特異結合する、単離された抗体を提供する。

【0141】

本発明は更に、（a）SEQ ID NO:22-42からなる群から選択したポリヌクレオチド配列を持つポリヌクレオチド、（b）SEQ ID NO:22-42からなる群から選択した或るポリヌクレオチド配列との少なくとも90%の同一性を有する或る天然ポリヌクレオチド配列を持つポリヌクレオチド、（c）（a）に相補的なポリヌクレオチド、（d）（b）に相補的なポリヌクレオチド、および（e）（a）～（d）のRNA等価物、からなる群から選択した

、単離されたポリヌクレオチドを提供する。一実施態様で該ポリヌクレオチドは、少なくとも60の連続したヌクレオチド群からなる。

【0142】

本発明は更に、サンプル中の標的ポリヌクレオチドを検出する一方法を提供する。ここで、標的ポリヌクレオチドは、(a) SEQ ID NO:22-42からなる群から選択した或るポリヌクレオチド配列を持つポリヌクレオチド、(b) SEQ ID NO:22-42からなる群から選択した或るポリヌクレオチド配列との少なくとも90%の同一性を有する天然ポリヌクレオチド配列を持つポリヌクレオチド、(c) (a)のポリヌクレオチドに相補的なポリヌクレオチド、(d) (b)のポリヌクレオチドに相補的なポリヌクレオチド、および(e) (a)～(d)のRNA等価物、からなる群から選択した、或るポリヌクレオチドの配列を持つ。10
検出方法は、(a) サンプル中の上記標的ポリヌクレオチドに相補的な或る配列からなる少なくとも20の連続したヌクレオチド群からなる或るプローブを用いて該サンプルをハイブリダイズする過程と、(b) 該ハイブリダイゼーション複合体の有無を検出し、複合体が存在すればオプションでその量を検出する過程からなる。該プローブと該標的ポリヌクレオチドあるいはその断片との間でハイブリダイゼーション複合体が形成されるような条件下で、プローブは、該標的ポリヌクレオチドに対し特異的にハイブリダイズする。一実施態様では、プローブは少なくとも60の連続したヌクレオチド群からなる。

【0143】

本発明はまた、サンプル中の標的ポリヌクレオチドを検出する方法を提供する。ここで、標的ポリヌクレオチドは、(a) SEQ ID NO:22-42からなる群から選択したポリヌクレオチド配列を含むポリヌクレオチド、(b) SEQ ID NO:22-42からなる群から選択したポリヌクレオチド配列と少なくとも90%の相同性を有する天然のポリヌクレオチド配列を含むポリヌクレオチド、(c) (a)のポリヌクレオチドに相補的なポリヌクレオチド、(d) (b)のポリヌクレオチドに相補的なポリヌクレオチド、および(e) (a)～(d)のRNA等価物、からなる群から選択された配列のポリヌクレオチドを有する。検出方法は、(a) ポリメラーゼ連鎖反応増幅を用いて標的ポリヌクレオチドまたはその断片を増幅する過程と、(b) 増幅した標的ポリヌクレオチドまたはその断片の有無を検出し、該標的ポリヌクレオチドまたはその断片が存在すればオプションでその量を検出する過程からなる。20

【0144】

本発明は更に、或る有効量のポリペプチドと薬物として許容し得る或る賦形剤とからなる、或る組成物を提供する。有効量のポリペプチドは、(a) SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチド、(b) SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列との少なくとも90%の同一性を有する天然のアミノ酸配列を含むポリペプチド、(c) SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を含むポリペプチドの生物学的活性断片、および(d) SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチドの免疫原性断片、からなる群から選択される。一実施態様で上記組成物は、SEQ ID NO:1-21からなる群から選択された或るアミノ酸配列を持つ。更に、本発明は、機能的CGDDの発現の低下を伴う疾患や症状の治療を要する患者に上記組成物を投与する方法を提供する。30

【0145】

本発明はまた、(a) SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチド、(b) SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列との少なくとも90%の同一性を有する天然アミノ酸配列を持つポリペプチド、(c) SEQ ID NO:1-21からなる群から選択したアミノ酸配列を持つポリペプチドの生物学的活性断片、および(d) SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチドの免疫原性断片、からなる群から選択したポリペプチドのアゴニストとしての有効性を確認するために、或る化合物をスクリーニングする方法を提供する。スクリーニング方法は、(a) 該ポリペプチドを有するサンプルを或る化合物に曝す過程と、(b) サンプル中のアゴニスト活性を検出する過程からなる。別法では、本発明は、この方法で同定したアゴニ 40 50

スト化合物と許容される医薬用賦形剤とを有する、或る組成物を提供する。更なる別法で本発明は、機能的CGDDの発現の低下を伴う疾患や症状の治療を要する患者への、この組成物の投与方法を提供する。

【0146】

本発明は更に、(a) SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチド、(b) SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列との少なくとも90%の同一性を有する天然アミノ酸配列を持つポリペプチド、(c) SEQ ID NO:1-21からなる群から選択したアミノ酸配列を持つポリペプチドの生物学的活性断片、および(d) SEQ ID NO:1-21からなる群から選択したアミノ酸配列を持つポリペプチドの免疫原性断片、からなる群から選択したポリペプチドのアンタゴニストとしての有効性につき、或る化合物をスクリーニングする方法を提供する。スクリーニング方法は、(a) 該ポリペプチドを含むサンプルを或る化合物に曝す過程と、(b) サンプル中のアンタゴニスト活性を検出する過程からなる。或いは本発明は、この方法で同定したアンタゴニスト化合物と薬物として許容し得る賦形剤とを有する組成物を提供する。更なる別法で本発明は、機能的CGDDの過剰発現を伴う疾患や症状の治療を要する患者への、この組成物の投与方法を提供する。

10

【0147】

本発明は更に、(a) SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチド、(b) SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列との少なくとも90%の同一性を有する或る天然アミノ酸配列を持つポリペプチド、(c) SEQ ID NO:1-21からなる群から選択したアミノ酸配列を持つポリペプチドの生物学的活性断片、または(d) SEQ ID NO:1-21からなる群から選択したアミノ酸配列を持つポリペプチドの免疫原性断片、からなる群から選択した或るポリペプチドに特異結合する或る化合物をスクリーニングする方法を提供する。スクリーニング方法は、(a) 該ポリペプチドを適切な条件下で少なくとも1つの試験化合物に混合させる過程と、(b) この試験化合物と該ポリペプチドとの結合を検出し、それにより該ポリペプチドに特異結合する化合物を同定する過程からなる。

20

【0148】

本発明は更に、(a) SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチド、(b) SEQ ID NO:1-21からなる群から選択したアミノ酸配列との少なくとも90%の同一性を有する或る天然アミノ酸配列を持つポリペプチド、(c) SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチドの生物学的活性断片、および(d) SEQ ID NO:1-21からなる群から選択した或るアミノ酸配列を持つポリペプチドの免疫原性断片、からなる群から選択した或るポリペプチドの活性をモジュレートする或る化合物をスクリーニングする方法を提供する。スクリーニング方法は、(a) 該ポリペプチドの活性にとり許容し得る条件下で、該ポリペプチドを少なくとも1つの試験化合物と混合する過程と、(b) 該ポリペプチドの活性をこの試験化合物の存在下で算定する過程と、(c) この試験化合物の存在下での該ポリペプチドの活性をこの試験化合物の不存在下での該ポリペプチドの活性と比較する過程からなり、この試験化合物の存在下での該ポリペプチドの活性の変化は、該ポリペプチドの活性をモジュレートする化合物を標示する。

30

40

【0149】

更に本発明は、SEQ ID NO:22-42からなる群から選択した或るポリヌクレオチド配列を持つ標的ポリヌクレオチドの発現を改変する効果につき、或る化合物をスクリーニングする方法を提供する。この方法は、(a) この標的ポリヌクレオチドを有するサンプルを或る化合物に曝露する過程と、(b) この標的ポリヌクレオチドの発現の改変を検出する過程と、(c) 可変量のこの化合物の存在下でのこの標的ポリヌクレオチドの発現と、この化合物の不存在下での発現とを比較する過程とからなる。

【0150】

本発明は更に、試験化合物の毒性の算定方法を提供する。この方法には、以下の過程があ

50

る。(a) 核酸群を有する生体サンプルを試験化合物で処理する過程。(b) 処理済み生体サンプルの核酸群をハイブリダイズする過程。この過程には、次のようなプローブを用いる。(i) SEQ ID NO:22-42からなる群から選択した或るポリヌクレオチド配列を持つポリヌクレオチド、(ii) SEQ ID NO:22-42からなる群から選択した或るポリヌクレオチド配列との少なくとも90%の同一性を有する天然ポリヌクレオチド配列を持つポリヌクレオチド、(iii) (i)に相補的な配列を持つポリヌクレオチド、(iv) (ii)のポリヌクレオチドに相補的なポリヌクレオチド、(v) (i)～(iv)のRNA等価物、からなる群から選択した或るポリヌクレオチドの少なくとも20の連続したヌクレオチド群からなるプローブである。ハイブリダイゼーションは、上記プローブと生体サンプル中の標的ポリヌクレオチドとの間に特異的ハイブリダイゼーション複合体が形成されるような条件下で生じる。上記標的ポリヌクレオチドは、(i) SEQ ID NO:22-42からなる群から選択した或るポリヌクレオチド配列を持つポリヌクレオチド、(ii) SEQ ID NO:22-42からなる群から選択した或るポリヌクレオチド配列との少なくとも90%の同一性を有する天然ポリヌクレオチド配列を持つポリヌクレオチド、(iii) (i)のポリヌクレオチドに相補的なポリヌクレオチド、(iv) (ii)のポリヌクレオチドに相補的なポリヌクレオチド、(v) (i)～(iv)のRNA等価物、からなる群から選択する。あるいは、標的ポリヌクレオチドは、上記(i)～(v)からなる群から選択したポリヌクレオチド配列の断片を持つ。毒性の算定方法には更に、以下の過程がある。(c) ハイブリダイゼーション複合体の量を定量する過程と、(d) 処理済み生体サンプル中のハイブリダイゼーション複合体の量を、非処理の生体サンプル中のハイブリダイゼーション複合体の量と比較する過程である。処理済み生体サンプル中のハイブリダイゼーション複合体の量の差異が、試験化合物の毒性を示す。

【発明を実施するための最良の形態】

【0151】

(本発明の記載について)

本発明のタンパク質、ヌクレオチド配列および方法について説明するが、その前に、説明した特定の装置、材料および方法に本発明が限定されるものではなく、修正され得ることを理解されたい。また、ここで使用する専門用語は特定の実施態様を説明する目的で用いたものに過ぎず、特許請求の範囲にのみ限定される本発明の範囲を限定することを意図したものではないことも併せて理解されたい。

【0152】

請求の範囲および明細書中で用いている単数形の「或る」および「その(この)」の表記は、文脈から明らかにそうでないとされる場合を除いて複数のものを指す場合もあることに注意されたい。したがって、例えば「或る宿主細胞」と記されている場合にはそのような宿主細胞が複数あることもあり、「或る抗体」と記されている場合には1個以上の抗体、および、当業者に公知の抗体の等価物などについても言及している。

【0153】

本明細書中で用いる全ての技術用語および科学用語は、特に定義されている場合を除き、当業者に一般に理解されている意味と同じ意味を有する。本明細書で説明するものと類似あるいは同等の任意の装置、材料および方法を用いて本発明の実施または試験を行うことができるが、ここでは好適な装置、材料、方法について説明する。本発明で言及する全ての刊行物は、刊行物中で報告されていて且つ本発明に関係して用い得る、細胞、プロトコル、試薬およびベクターについて説明および開示する目的で引用しているものである。本明細書のいかなる開示内容も、本発明が先行技術の効力によってこのような開示に対して先行する権利を与えられていないことを認めるものではない。

【0154】

(定義)

用語「CGDD」は、天然、合成、半合成或いは組換え体などの、全ての種(特にウシ、ヒツジ、ブタ、ネズミ、ウマ及びヒトを含む哺乳動物)から得られる実質的に精製されたCGDDのアミノ酸配列を指す。

【0155】

用語「アゴニスト」は、CGDDの生物学的活性を強化、或いは模倣する分子を指す。アゴニストの例として、タンパク質、核酸、糖質、小分子その他の任意の化合物や組成物を挙げることができるが、これらはCGDDと直接相互作用することによって、或いはCGDDが関与する生物学的経路の構成成分に作用することによって、CGDDの活性を調節する。

【0156】

「対立遺伝子変異体」は、CGDDをコードする遺伝子の、別の形態である。対立遺伝子変異体群は、核酸配列における少なくとも1つの突然変異から生じ得る他、変容したmRNA群を生じ得る。また、生じ得るポリペプチドの構造または機能は、変容することもしないこともある。或る遺伝子は、その天然型の対立遺伝子変異体を全く持たない場合もあり、1個以上持つこともある。対立遺伝子変異体を生じさせる通常の突然変異性変化は一般に、ヌクレオチドの自然な欠失、付加または置換による。これら各種の変化は、単独であるいは他の変化と共に、或る配列内で1回以上、生じ得る。

10

【0157】

CGDDをコードする「変容した／改変された」核酸配列としては、様々なヌクレオチドの欠失、挿入、あるいは置換がありながら、CGDDと同じポリペプチド、あるいはCGDDの機能特性の少なくとも1つを備えるポリペプチドをもたらし配列もある。この定義には、CGDDをコードするポリヌクレオチド配列にとり正常な染色体の遺伝子座ではない位置での、対立遺伝子変異体群への不適当或いは予期しないハイブリダイゼーションを含み、また、CGDDをコードするポリヌクレオチドの或る特定オリゴヌクレオチドプローブを用いて容易に検出可能な或いは検出困難な多型性を含む。コードされるタンパク質も「変容する／改変される」ことがあり、また、サイレント変化を生じた結果、機能的に等価なCGDDとなるような、アミノ酸残基の欠失、挿入または置換を持ち得る。意図的なアミノ酸置換は、生物学的あるいは免疫学的にCGDDの活性が保持される範囲で、残基群の、極性、電荷、溶解度、疎水性、親水性、および／または両親媒性についての、類似性に基づいて成され得る。例えば、負に帯電したアミノ酸にはアスパラギン酸およびグルタミン酸があり、正に帯電したアミノ酸にはリジンおよびアルギニンがある。親水性値が近似した非荷電極性側鎖を持つアミノ酸としては、アスパラギンとグルタミン、およびセリンとトレオニンを含みうる。親水性値が近似した非荷電側鎖を持つアミノ酸としては、ロイシンとイソロイシンとバリン、グリシンとアラニン、およびフェニルアラニンとチロシンを含みうる。

20

30

【0158】

用語「アミノ酸」および「アミノ酸配列」は、オリゴペプチド、ペプチド、ポリペプチド、タンパク質配列、あるいはそれらのいずれかの断片を指し、天然分子または合成分子を指す。「アミノ酸配列」が或る天然タンパク質分子の配列を指す場合、「アミノ酸配列」および類似の用語は、記載したそのタンパク質分子に関連する完全で元のままのアミノ酸配列に限定するものではない。

【0159】

「増幅」は、或る核酸配列の付加的複製物を作製する行為に関する。増幅には通常、当業者に公知のポリメラーゼ連鎖反応（PCR）技術を用いる。

【0160】

用語「アンタゴニスト」は、CGDDの生物学的活性を阻害あるいは減弱する分子を指す。アンタゴニストとしては、CGDDと直接相互作用すること、あるいはCGDDが関与する生物学的経路の各成分に作用することによってCGDDの活性をモジュレートする、抗体などタンパク質、核酸、糖質、小分子、または任意の他の化合物や組成物があり得る。

40

【0161】

用語「抗体」は、エピトープの決定基と結合できる、無傷の免疫グロブリン分子やそれらの断片、例えばFab、F(ab')₂、およびFv断片を指す。CGDDポリペプチド群と結合する抗体類の作製には、免疫抗原として、無傷ポリペプチド群を用いることができ、または、当該の小ペプチド群を有する断片群を用い得る。マウス、ラット、ウサギなどの動物を免疫化するために用いるポリペプチドまたはオリゴペプチドの由来は、RNAの翻訳の場合や、

50

または化学合成があり得る。また、それらは所望により、キャリアタンパク質に抱合し得る。通常用いられるキャリアであってペプチドと化学結合するキャリアの例は、ウシ血清アルブミン、サイログロブリン、および、スカシガイのヘモシアニン (KLH) などがある。結合したこのペプチドは、前記動物を免疫化するために用いる。

【0162】

用語「抗原決定基」は、特定の抗体と接触する、分子の領域（すなわちエピトープ）を指す。タンパク質またはタンパク質断片を用いて宿主動物を免疫化する場合、タンパク質の多数の領域が、抗原決定基（タンパク質の特定の領域または3次元構造）に特異結合する抗体の産生を誘発し得る。或る抗原決定基は、或る抗体への結合について、無傷抗原（すなわち免疫応答を引き出すために用いられる免疫原）と競合し得る。

10

【0163】

用語「アプタマー (aptamer)」は、核酸またはオリゴヌクレオチド分子であって、特定の分子ターゲットに結合する分子を指す。アプタマーは *in vitro* での進化プロセスに由来する（例えば、SELEX (Systematic Evolution of Ligands by EXponential Enrichment) の略、試験管内選択法）、米国特許第5,270,163号に記述）。これは、大規模な組合せライブラリ群から標的特異的アプタマー配列を選択するプロセスである。アプタマーの構成は二本鎖または一本鎖であり、デオキシリボヌクレオチド、リボヌクレオチド、ヌクレオチド誘導体または他のヌクレオチド様分子を含み得る。アプタマーの各ヌクレオチド成分は修飾された糖基を持つことがあり（例えば、リボヌクレオチドの2'-OH基は2'-Fまたは2'-NH₂によって置換され得る）、これは、ヌクレアーゼへの抵抗性または血中でのより長い寿命など、所望の特性を向上し得る。アプタマーを他の分子（例えば高分子量のキャリア）と抱合させることにより、循環系からのこのアプタマーのクリアランスを遅らせ得る。アプタマー類は、例えば或る架橋剤の光活性化によって、それらの同種リガンド群と特異的に架橋させ得る（例えば、Brody, E.N.およびL. Gold (2000) J. Biotechnol. 74:5-13を参照）。

20

【0164】

用語「イントラマー (intramer)」は、*in vivo* で発現されるアプタマーを指す。例えば、ワクシニアウイルスに基づく或るRNA発現系を用いて、白血球の細胞質内で特定のRNAアプタマー類が高レベルに発現されている (Blind, M.他(1999) Proc.Natl Acad. Sci. USA 96:3606-3610)。

30

【0165】

用語「スピーゲルマー (spiegelmer)」は、アプタマーの内、L-DNA、L-RNAなどの左旋性ヌクレオチド誘導体または左旋性ヌクレオチド様分子などを指す。左旋性のヌクレオチド群を持つアプタマー類は、右旋性ヌクレオチド群を持つ基質に通常は作用する、天然酵素類による分解に対して耐性がある。

【0166】

用語「アンチセンス」は、或る特定の核酸配列の「センス」（コーディング）鎖との塩基対を形成し得る任意の組成物を指す。アンチセンス組成物としては、DNAや、RNAや、ペプチド核酸 (PNA) や、修飾されたバックボーン連結たとえばホスホロチオ酸、メチルホスホン酸またはベンジルホスホン酸などを有するオリゴヌクレオチドや、修飾された糖基たとえば2'-メトキシエチル糖または2'-メトキシエトキシ糖などを有するオリゴヌクレオチドや、あるいは修飾された塩基たとえば5-メチルシトシン、2-デオキシウラシルまたは7-デアザ-2'-デオキシグアノシンなどを有するオリゴヌクレオチドがあり得る。アンチセンス分子は、化学合成または転写など、任意の方法で製造することができる。相補的アンチセンス分子は、細胞に導入されると、細胞が産生した天然核酸配列との塩基対を形成し、二重鎖を形成して転写または翻訳を妨害する。「負」または「マイナス」という表現は、或る参考DNA分子のアンチセンス鎖を指すことがあり、「正」または「プラス」という表現は、或る参考DNA分子のセンス鎖を意味し得る。

40

【0167】

用語「生物学的に活性」は、或る天然分子の構造的、調節的、あるいは生化学的な機能を

50

有するタンパク質を指す。同様に、用語「免疫学的に活性」または「免疫原性」は、天然、組換え、または合成CGDD、またはそれらの任意のオリゴペプチドが、適当な動物或いは細胞内の特定の免疫応答を誘発して特定の抗体と結合する能力を指す。

【0168】

「相補」は、塩基対合によってアニールする2つの一本鎖核酸配列間の関係を指す。例えば、「5'-AGT-3'」は、その相補配列「3'-TCA-5'」との対を形成する。

【0169】

「～のポリヌクレオチド配列を含む（持つ）組成物」または「～のアミノ酸配列を含む（持つ）組成物」は、広い意味で、指定のポリヌクレオチド配列若しくはアミノ酸配列を持つ、任意の組成物を指す。この組成物には、乾燥製剤または水溶液が含まれ得る。CGDDをコード、若しくはその断片をコードするポリヌクレオチド配列を持つ組成物は、ハイブリダイゼーションプローブとして使用され得る。これらプローブは、凍結乾燥形態で貯蔵でき、また、糖質などの安定化剤と結合させ得る。ハイブリダイゼーションにおいては、塩（例えばNaCl）、界面活性剤（例えばドデシル硫酸ナトリウム；SDS）および他の成分（例えばデンハート液、粉乳、サケ精子DNAなど）を有する水溶液中に、プローブを分散させ得る。

10

【0170】

「コンセンサス配列」は、不要な塩基を除くためにDNA配列解析を繰り返し行い、XL-PCRキット（Applied Biosystems, Foster City CA）を用いて5'および／または3'の方向に伸長され、再度シーケンシングされた核酸配列、またはGELVIEWフラグメント結合システム（GCG, Madison, WI）またはPhrap（University of Washington, Seattle WA）などのフラグメント結合用コンピュータプログラムを用いて1つ以上のオーバーラップするcDNAやEST、またはゲノムDNA断片から構築（アセンブリ）された核酸配列を指す。伸長および構築の両方を行ってコンセンサス配列を産生する配列もある。

20

【0171】

用語「保存的なアミノ酸置換」は、元のタンパク質の特性を殆ど変えないと予測される置換を指す。すなわち、そのタンパク質の構造、特にその機能が保存され、置換によって大きくは変わらない。下表は、タンパク質内で元のアミノ酸と置換され得るアミノ酸であり、保存的なアミノ酸置換と認められるアミノ酸を示す。

【0172】

30

元の残基	保存的な置換	
Ala	Gly, Ser	
Arg	His, Lys	
Asn	Asp, Gln, His	
Asp	Asn, Glu	
Cys	Ala, Ser	
Gln	Asn, Glu, His	
Glu	Asp, Gln, His	
Gly	Ala	
His	Asn, Arg, Gln, Glu	10
Ile	Leu, Val	
Leu	Ile, Val	
Lys	Arg, Gln, Glu	
Met	Leu, Ile	
Phe	His, Met, Leu, Trp, Tyr	
Ser	Cys, Thr	
Thr	Ser, Val	
Trp	Phe, Tyr	
Tyr	His, Phe, Trp	
Val	Ile, Leu, Thr	20

【0173】

保存的なアミノ酸置換では通常、(a)置換領域におけるポリペプチドのバックボーン構造、例えばβシートやα螺旋構造、(b)置換部位における分子の電荷または疎水性、および/または(c)側鎖の大部分、を保持する。

【0174】

用語「欠失」は、1個以上のアミノ酸残基が欠如するアミノ酸配列内の変化、あるいは1個以上のヌクレオチドが欠如するヌクレオチド配列内の変化を指す。

【0175】

用語「誘導体」は、化学修飾されたポリヌクレオチドまたはポリペプチドを指す。例えば、アルキル基、アシル基、ヒドロキシル基またはアミノ基による水素の置換は、ポリヌクレオチドの化学修飾に含まれ得る。ポリヌクレオチド誘導体は、天然分子の生物学的または免疫学的機能を少なくとも1つは保持しているポリペプチドをコードする。ポリペプチド誘導体は、次のようなプロセスによって修飾されたポリペプチドである。すなわち、グリコシル化、ポリエチレングリコール化(pegylation)、あるいは任意の同様なプロセスであって、誘導元のポリペプチドの少なくとも1つの生物学的若しくは免疫学的機能を保持するプロセスである。

【0176】

「検出可能な標識」は、測定可能な信号を発生し得る、ポリヌクレオチドやポリペプチドに共有結合あるいは非共有結合するレポーター分子や酵素を指す。 40

【0177】

「差次的発現」は、少なくとも2例のサンプルを比較して判定する、増加(上方調節)、あるいは減少(下方調節)、または欠損した、遺伝子またはタンパク質の発現を指す。このような比較は例えば、処理済サンプルと不処理サンプル、または病態サンプルと健常サンプルとの間で行われ得る。

【0178】

「エキソンシャッフリング」は、異なるコード領域(エキソン)群の組換えを意味する。或るエキソンは、コードするタンパク質の1つの構造的または機能的ドメインを代表し得るため、安定したサブストラクチャー群の新たな組合せによって新規のタンパク質群を構 50

築することが可能であり、新たなタンパク質機能の進化を促進できる。

【0179】

「断片」は、CGDDの又はCGDDをコードするポリヌクレオチドの固有の部分であって、その親配列 (parent sequence) と配列は同一であるが親配列より長さが短いものを指す。或る断片は、定義された配列の全長から1ヌクレオチド／アミノ酸残基を差し引いた長さよりも短い長さを有し得る。例えば或る断片は、5～1000の連続したヌクレオチドまたはアミノ酸残基を有し得る。プローブ、プライマー、抗原、治療用分子として、あるいはその他の目的のために用いられる或る断片は、少なくとも長さ5、10、15、20、25、30、40、50、60、75、100、150、250若しくは500の、連続したヌクレオチドあるいはアミノ酸残基であり得る。断片は、或る分子の特定領域から優先的に選択し得る。例えば、或るポリペプチド断片は、定義された或る配列内に見られるような或るポリペプチドの最初の250または500アミノ酸（または最初の25%または50%）から選択した、或る長さの連続したアミノ酸を持ち得る。これらの長さは明らかに例として挙げているものであり、本発明の実施態様では、配列表、表および図面を含む本明細書が支持する任意の長さであり得る。

【0180】

SEQ ID NO:22-42の或る断片は、或る固有のポリヌクレオチド配列領域を持つ。この領域は、SEQ ID NO:22-42を特異的に同定するものであり、例えばこの断片を得たゲノム中の他のいかなる配列とも異なる。SEQ ID NO:22-42の或る断片は、例えば、ハイブリダイゼーションや増幅技術に、またはSEQ ID NO:22-42を関連ポリヌクレオチド配列から区別する類似の方法に有用である。SEQ ID NO:22-42の或る断片の正確な長さは、また、その断片が対応するSEQ ID NO:22-42の領域は、その断片に意図した目的に基づき当業者が慣例的に決定し得る。

【0181】

SEQ ID NO:1-21の或る断片は、SEQ ID NO:22-42の或る断片によってコードされる。SEQ ID NO:1-21の或る断片は、SEQ ID NO:1-21を特異的に同定する固有のアミノ酸配列の領域を持つ。例えば、SEQ ID NO:1-21の或る断片は、SEQ ID NO:1-21を特異的に認識する抗体の開発における免疫原性ペプチドとして有用である。SEQ ID NO:1-21の或る断片の正確な長さは、また、この断片に対応するSEQ ID NO:1-21の領域は、その断片に意図する目的に基づき、当業者が慣例的に判定し得る。

【0182】

「完全長」ポリヌクレオチド配列とは、少なくとも1つの翻訳開始コドン（例えばメチオニン）と、それに続く1オープンリーディングフレームおよび翻訳終止コドンを含む配列である。或る「完全長」ポリヌクレオチド配列は、或る「完全長」ポリペプチド配列をコードする。

【0183】

「相同性」の語は、2つ以上のポリヌクレオチド配列または2つ以上のポリペプチド配列の、配列類似性、互換性、または配列同一性を意味する。

【0184】

ポリヌクレオチド配列に適用される「一致率」および「～%同一」の語は、標準化されたアルゴリズムを用いてアラインメントされた、2つ以上のポリヌクレオチド配列間で一致する残基の割合を意味する。標準化アルゴリズムは、2配列間のアラインメントを最適化するため、標準化された再現性のある方法で比較対象の2配列内にギャップ群を挿入し得るので、2つの配列をより有意に比較できる。

【0185】

ポリヌクレオチド配列間の一致率を判定するには、MEGALIGN version 3.12e配列アラインメントプログラムなどに組込まれているCLUSTAL Vアルゴリズムの、デフォルトのパラメータ群を用い得る。このプログラムは、LASERGENE ソフトウェアパッケージ（一組の分子生物学的分析プログラム）（DNASTAR, Madison WI）の一部である。このCLUSTAL Vは、Higgins, D.G.およびP.M. Sharp (1989) CABIOS 5:151-153、Higgins, D.G. 他 (1992) CAB

IOS 8:189-191に記載されている。ポリヌクレオチド配列をペアワイズアラインメントする際のデフォルトパラメータは、Ktuple=2、gap penalty=5、window=4、「diagonals saved」=4と設定される。「weighted」残基重み付け表がデフォルトで選択される。CLUSTAL Vによる一致率の報告は、アラインメントされたポリヌクレオチド配列間の「percent similarity (類似性パーセント)」としてなされる。

【0186】

あるいは、米国国立バイオテクノロジー情報センター (NCBI) の Basic Local Alignment Search Tool (BLAST) が、一般的に用いられ、且つ、無料で利用可能な配列比較アルゴリズム一式を提供している (Altschul, S.F. 他 (1990) J. Mol. Biol. 215:403-410)。BLASTアルゴリズムは、幾つかの情報源から入手可能であり、メリーランド州ベセスダにある NCBI および インターネット (<http://www.ncbi.nlm.nih.gov/BLAST/>) から入手可能である。この BLAST ソフトウェア一式には、既知のポリヌクレオチド配列と、様々なデータベースの別のポリヌクレオチド配列とのアラインメントに用いる「blastn」を含む、様々な配列分析プログラムが含まれる。「BLAST 2 Sequences」と呼ばれるツールも入手可能であり、これは2つのヌクレオチド配列を直接のペアワイズで比較するために用いられる。「BLAST 2 Sequences」は、<http://www.ncbi.nlm.nih.gov/gorf/bl2.html> にアクセスして、対話形式で利用できる。「BLAST 2 Sequences」ツールは、blastn および blastp (以下に記載) の両方に用い得る。BLAST プログラムは、一般的には、ギャップ (gap) などのパラメータをデフォルト設定にセットして用いる。例えば、2つのヌクレオチド配列を比較するには、デフォルトパラメータに設定した「BLAST 2 Sequences」ツール Version 2.0.12 (2000年4月21日) を用いて blastn を実行し得る。デフォルトパラメータの設定例を以下に示す。

【0187】

Matrix: BLOSUM62
Reward for match: 1
Penalty for mismatch: -2
Open Gap: 5 and Extension Gap: 2 penalties
Gap x drop-off: 50
Expect: 10
Word Size: 11
Filter: on

【0188】

一致率は、ある定義された配列の全長 (例えば特定の SEQ ID ナンバーで定義された配列) について測定し得る。あるいは、より短い長さ、例えば、定義された、より大きな配列から得た断片 (例えば少なくとも 20、30、40、50、70、100 または少なくとも 200 の連続したヌクレオチドの断片) の長さの一致率も測定し得る。ここに挙げた長さは単なる例示的なものに過ぎず、表、図および配列リストを含めた本明細書に記載された配列が支持する任意の断片長を用いて、一致率を測定し得る或る長さを説明し得ることを理解されたい。

【0189】

高度の同一性を示さない核酸配列が、それにもかかわらず遺伝子コードの縮重が原因で、類似のアミノ酸配列をコードする場合がある。この縮重を利用して核酸配列内で変化を生じさせて、全ての核酸配列が実質上同一のタンパク質をコードするような多数の核酸配列を生成し得ることを理解されたい。

【0190】

ポリペプチド配列に用いられる用語「一致率」または「～% 同一」とは、標準化されたアルゴリズムを用いてアラインメントされる2つ以上のポリペプチド配列間の一致する残基の百分率のことである。ポリペプチド配列アラインメントの方法は公知である。保存的アミノ酸置換を考慮するアラインメント方法もある。既に詳述したこのような保存的置換は通常、置換部位の電荷および疎水性を保存するので、ポリペプチドの構造を (したがって

10

20

30

40

50

機能も)保存する。

【0191】

ポリペプチド配列間の一致率は、MEGALIGN version 3.12e配列アラインメントプログラムに組み込まれているようなCLUSTAL Vアルゴリズムのデフォルトのパラメータ群を用いて判定できる(参照先と共に上述)。ポリペプチド配列をCLUSTAL Vを用いてペアワイズアラインメントする際のデフォルトパラメータは、Ktuple=1、gap penalty=3、window=5、「diagonals saved」=5と設定される。PAM250マトリクスがデフォルト残基重み付け表として選択される。ポリヌクレオチドのアラインメントと同様に、アラインメントされたポリペプチド配列の対の一致率は、CLUSTAL Vによって「percent similarity(類似性パーセント)」として報告される。

10

【0192】

あるいは、NCBI BLASTソフトウェア一式を用い得る。例えば、2つのポリペプチド配列をペアワイズで比較する場合、「BLAST 2 Sequences」ツールVersion 2.0.12(2000年4月21日)のblastpをデフォルトパラメータに設定して用い得る。デフォルトパラメータの設定例を以下に示す。

【0193】

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

20

【0194】

一致率は、ある定義されたポリペプチド配列の全長(例えば特定のSEQ IDナンバーで定義された配列)について測定し得る。あるいは、より短い長さ、例えば、定義された、より大きなポリペプチド配列から得た断片(例えば少なくとも15、20、30、40、50、70、または少なくとも150の連続した残基の断片)の長さの一致率も測定し得る。ここに挙げた長さは単なる例示的なものに過ぎず、表、図および配列リストを含めた本明細書に記載された配列が支持する任意の断片長を用いて、一致率を測定し得る或る長さを説明し得ることを理解されたい。

30

【0195】

「ヒト人工染色体(HAC)」は直鎖状の微小染色体であり、約6 kb ~ 10 MbのサイズのDNA配列を持ち得る。また、染色体の複製、分離および維持に必要な、全てのエレメントを持つ。

【0196】

用語「ヒト化抗体」は、もとの結合能力を保持しつつ、よりヒトの抗体に似せるために、非抗原結合領域のアミノ酸配列を改変した抗体分子を指す。

【0197】

「ハイブリダイゼーション」とは、所定のハイブリダイゼーション条件下で、或る一本鎖ポリヌクレオチドが或る相補的な一本鎖と塩基対を形成するアニーリングのプロセスである。特異的ハイブリダイゼーションは、2つの核酸配列が高い相補性を共有することの指標である。特異的ハイブリダイゼーション複合体は許容されるアニーリング条件下で形成され、1回以上の「洗浄」ステップ後もハイブリダイズされたままである。洗浄ステップは、ハイブリダイゼーションプロセスのストリンジェンシーを決定する際に特に重要であり、よりストリンジェントな条件では、非特異結合(すなわち完全には一致しない核酸鎖対間の結合)が減少する。核酸配列のアニーリングに関する許容条件は、当業者が慣例的に決定できる。アニール条件はどのハイブリダイゼーション実験でも一定であり得るが、洗浄条件は、所望のストリンジェンシー、したがってハイブリダイゼーション特異性を得るように、実験ごとに変更し得る。許容的アニーリング条件は、例えば、温度が68℃で、約6×SSC、約1%(w/v)のSDS、並びに約100 μg/mlの、せん断して変性した

40

50

サケ精子DNAの存在下である。

【0198】

一般に、ハイブリダイゼーションのストリンジェンシーは或る程度、洗浄ステップを実行する温度を基準にして表すことができる。このような洗浄温度は通常、所定のイオン強度およびpHにおける特定配列の融点 (T_m) より約 $5 \sim 20^\circ\text{C}$ 低くなるように選択する。この T_m は、所定のイオン強度およびpHの条件下で、完全一致プローブに標的配列の50%がハイブリダイズする温度である。 T_m を計算する式および核酸のハイブリダイゼーション条件はよく知られており、Sambrook, J. 他(1989) Molecular Cloning: A Laboratory Manual, 第2版, 1-3巻, Cold Spring Harbor Press, Plainview NYに記載されており、特に2巻の9章を参照されたい。

10

【0199】

本発明のポリヌクレオチド間の高ストリンジェンシー条件のハイブリダイゼーションには、約 $0.2 \times \text{SSC}$ および約 0.1% のSDSの存在下、 68°C で1時間の洗浄条件を含む。あるいは、温度は約 65°C 、 60°C 、 55°C 、または 42°C を用い得る。SSC濃度は、約 0.1% のSDS存在下で、約 $0.1 \sim 2 \times \text{SSC}$ の範囲で変更し得る。通常は、ブロッキング剤を用いて非特異ハイブリダイゼーションを阻止する。このようなブロッキング剤には、例えば、約 $100 \sim 200 \mu\text{g/ml}$ の、せん断した変性サケ精子DNAがある。特定条件下で、例えばRNAとDNAのハイブリダイゼーションでは、有機溶剤、例えば約 $35 \sim 50\% \text{ v/v}$ の濃度のホルムアミドを用いることもできる。洗浄条件の有用なバリエーションは、当業者には自明であろう。ハイブリダイゼーションは、特に高ストリンジェント条件下では、ヌクレオチド間の進化的な類似性を示唆し得る。進化的類似性は、ヌクレオチド群、およびヌクレオチドがコードするポリペプチド群について、或る同様の役割を強く示唆する。

20

【0200】

用語「ハイブリダイゼーション複合体」は、相補的な塩基間の水素結合の形成によって形成された、2つの核酸配列の複合体を指す。ハイブリダイゼーション複合体は、溶解状態で形成し得る (C_0t または R_0t 解析など)。あるいは、一方の核酸配列が溶解状態で存在し、もう一方の核酸配列が固体支持体 (例えば紙、膜、フィルタ、チップ、ピンまたはガラススライド、あるいは他の適切な基板であって細胞若しくはその核酸が固定される基板) に固定されているような2つの核酸配列間に形成され得る。

【0201】

用語「挿入」あるいは「付加」は、1個以上のアミノ酸残基あるいはヌクレオチドがそれぞれ追加される、アミノ酸配列あるいは核酸配列における変化を指す。

30

【0202】

「免疫応答」は、炎症、外傷、免疫異常症、伝染性疾患または遺伝性疾患などに関連する症状を指し得る。これらの症状は、細胞および全身の防御系に作用し得る種々の因子、例えばサイトカイン、ケモカイン、その他のシグナル伝達分子の発現によって特徴づけ得る。

【0203】

用語「免疫原性断片」は、生物 (例えば哺乳類) に導入すると免疫応答を引き起こし得る、CGDDのポリペプチド断片またはオリゴペプチド断片を指す。用語「免疫原性断片」はまた、本明細書で開示するまたは当分野で周知のあらゆる抗体生産方法に有用な、CGDDの任意のポリペプチド断片またはオリゴペプチド断片をも指す。

40

【0204】

用語「マイクロアレイ」は、或る基板上の複数のポリヌクレオチド、ポリペプチドまたはその他の化合物の構成を指す。

【0205】

用語「エレメント」または「アレイエレメント」は、マイクロアレイ上に固有の指定された位置を有する、ポリヌクレオチド、ポリペプチドまたはその他の化合物を指す。

【0206】

用語「モジュレート」または「活性を調節」は、CGDDの活性を変化させることを指す。例

50

えば、モジュレートによって、CGDDのタンパク質活性の増減が、あるいは、結合特性またはその他の、生物学的特性、機能的特性あるいは免疫学的特性の変化が生じ得る。

【0207】

「核酸」および「核酸配列」の語は、ヌクレオチド、オリゴヌクレオチド、ポリヌクレオチドまたはこれらの断片を指す。「核酸」および「核酸配列」の語はまた、ゲノム起源または合成起源のDNAまたはRNAであって一本鎖または二本鎖であるかあるいはセンス鎖またはアンチセンス鎖を表し得るようなDNAまたはRNAや、ペプチド核酸（PNA）や、任意のDNA様またはRNA様物質を指すこともある。

【0208】

「機能的に連結した」は、第1の核酸配列と第2の核酸配列が機能的な関係にある状態を指す。例えば、或るプロモーターが或るコード配列の転写または発現に影響を及ぼす場合には、そのプロモーターはそのコード配列に機能的に連結している。機能的に連結したDNA配列群は非常に近接するか連続的に隣接することがあり、また、2つのタンパク質コード領域を結合するために必要な場合は同一リーディングフレーム内にあり得る。

【0209】

「ペプチド核酸（PNA）」は、末端がリジンで終わるアミノ酸残基のペプチドのバックボーンに結合した、少なくとも約5ヌクレオチドの長さのオリゴヌクレオチドを持つ、アンチセンス分子または抗遺伝子剤を指す。末端のリジンは、この組成に溶解性を与える。PNAは、相補的一本鎖DNAまたはRNAに優先的に結合して転写の伸長を停止させる。ポリエチレングリコール化することにより、細胞におけるPNAの寿命を延長し得る。

【0210】

CGDDの「翻訳後修飾」には、脂質化、グリコシル化、リン酸化、アセチル化、ラセミ化、タンパク質分解切断およびその他の当分野で既知の修飾を含み得る。これらのプロセスは、合成により、あるいは生化学的に生じ得る。生化学的修飾は、CGDDの酵素環境に依存し、細胞の種類によって異なることとなる。

【0211】

「プローブ」とは、同一配列あるいは対立遺伝子核酸配列、関連する核酸配列の検出に用いる、CGDDやそれらの相補配列、またはそれらの断片をコードする核酸配列のことである。プローブは、単離されたオリゴヌクレオチドまたはポリヌクレオチドであって、検出可能な標識またはレポーター分子に接着した配列である。典型的な標識には、放射性アイソトープ、リガンド、化学発光試薬および酵素がある。「プライマー」とは、相補的な塩基対を形成して標的ポリヌクレオチドにアニーリング可能な、通常はDNAオリゴヌクレオチドである短い核酸である。プライマーは次に、DNAポリメラーゼ酵素により、標的DNA鎖に沿って伸長され得る。プライマー対は、例えばポリメラーゼ連鎖反応（PCR）による、核酸配列の増幅（および同定）に用い得る。

【0212】

本発明に用いるプローブおよびプライマーは通常、既知の配列の、少なくとも15の連続したヌクレオチド群からなる。特異性を高めるため、長めのプローブおよびプライマー、例えば開示した核酸配列の少なくとも20、25、30、40、50、60、70、80、90、100または少なくとも150の連続したヌクレオチドからなるようなプローブおよびプライマーも用い得る。これよりもかなり長いプローブおよびプライマーもある。表、図面および配列リストを含む本明細書が支持する、任意の長さのヌクレオチドを用い得るものと理解されたい。

【0213】

プローブおよびプライマーの調製および使用方法については、Sambrook, J. 他(1989) Molecular Cloning: A Laboratory Manual, 第2版, 1-3巻, Cold Spring Harbor Press, Plainview NY、Ausubel, F.M. 他(1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY、Innis, M. 他(1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CAなどを参照されたい。PCRプライマー対を既知の配列から得るには、例えば、そのためのコンピュ

ータプログラム、例えばPrimer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA) を用い得る。

【0214】

プライマーとして用いるオリゴヌクレオチドの選択は、そのような目的のために本技術分野でよく知られているソフトウェアを用いて行う。例えばOLIGO 4.06ソフトウェアは、各100ヌクレオチドまでのPCRプライマー対の選択に有用であり、オリゴヌクレオチドおよび、最大5,000までの大きめのポリヌクレオチドであって32キロベースまでのインプットポリヌクレオチド配列から得た配列を分析するのにも有用である。類似のプライマー選択プログラムには、拡張能力のための追加機能が組込まれている。例えば、PrimOUプライマー選択プログラム（テキサス州ダラスにあるテキサス大学南西部医療センターのゲノムセンターから一般向けに入手可能）は、メガベース配列から特定のプライマーを選択することが可能であり、したがってゲノム全体の範囲でプライマーを設計するのに有用である。Primer3プライマー選択プログラム（Whitehead Institute/MIT Center for Genome Research（マサチューセッツ州ケンブリッジ）より入手可能）を用いれば、プライマー結合部位として避けたい配列を指定できる「非プライミングライブラリ（mispriming library）」を入力できる。Primer3は特に、マイクロアレイのためのオリゴヌクレオチドの選択に有用である（後二者のプライマー選択プログラムのソースコードは、それぞれの情報源から得てユーザー固有のニーズを満たすように修正し得る）。PrimerGenプログラム（英国ケンブリッジ市の英国ヒトゲノムマッピングプロジェクト-リソースセンターから一般向けに入手可能）は、多数の配列アラインメントに基づいてプライマーを設計し、それによって、アラインメントされた核酸配列の最大保存領域または最小保存領域のいずれかとハイブリダイズするようなプライマーの選択を可能にする。したがって、このプログラムは、独自のものであれ保存されたものであれ、オリゴヌクレオチドとポリヌクレオチド断片との同定に有用である。上記選択方法のいずれかによって同定したオリゴヌクレオチドおよびポリヌクレオチド断片は、ハイブリダイゼーション技術において、例えばPCRまたはシーケンシングプライマーとして、マイクロアレイエレメントとして、あるいは核酸のサンプルにおいて完全または部分的相補的ポリヌクレオチドを同定する特異プローブとして有用である。オリゴヌクレオチドの選択方法は、上記の方法に限定されるものではない。

10

20

30

【0215】

「組換え核酸」は、天然の配列ではなく、配列の、2つ以上の離れたセグメントを人工的に組合せた配列である。この人為的組合せはしばしば化学合成によって達成するが、より一般的には核酸の単離セグメントの人為的操作によって、例えばSambrookの文献（前出）に記載されているような遺伝子工学的手法によって達成する。組換え核酸の語は、単に核酸の一部が付加、置換または欠失により改変された核酸も含む。しばしば組換え核酸には、プロモーター配列に機能的に連結した核酸配列が含まれる。このような組換え核酸は、例えばある細胞を形質転換するために使用されるベクターの一部と成し得る。

【0216】

あるいはこのような組換え核酸は、ウイルスベクターの一部と成すことができ、ベクターは例えばワクシニアウイルスに基づくものであり得る。そのようなベクターは哺乳類に接種され、その組換え核酸が発現されて、その哺乳類内で防御免疫応答を誘導するように使用することができる。

40

【0217】

「調節エレメント」は、或る遺伝子の非翻訳領域に通常は由来する核酸配列であり、エンハンサー、プロモーター、イントロンおよび5'および3'の非翻訳領域（UTR）を含む。調節エレメントは、転写、翻訳またはRNA安定性を制御する宿主タンパク質またはウイルスタンパク質と相互作用する。

【0218】

「レポーター分子」は、核酸、アミノ酸または抗体の標識に用いる、化学的または生化学的な成分である。レポーター分子には、放射性核種、酵素、蛍光剤、化学発光剤、発色剤

50

、基質、補助因子、阻害因子、磁気粒子およびその他の当分野で既知の成分がある。

【0219】

或るDNA配列に対する「RNA等価物」は、基準となるDNA配列と同じ直鎖のヌクレオチド配列からなるが、生じる全ての窒素性塩基のチミンがウラシルに置換され、糖鎖のバックボーンがデオキシリボースではなくリボースからなる。

【0220】

用語「サンプル」は、その最も広い意味で用いられている。CGDD、CGDDをコードする核酸群、またはその断片群を含むと推定されるサンプルとしては、体液と、細胞や細胞から単離した染色体や細胞内小器官（オルガネラ）や膜からの抽出物と、細胞と、溶液中に存在するまたは基板に固定されたゲノムDNA、RNA、cDNAと、組織と、組織プリントなどがあり得る。 10

【0221】

用語「特異結合」および「特異的に結合する」は、タンパク質若しくはペプチドの、アゴニスト、抗体、アンタゴニスト、小分子、若しくは任意の天然若しくは合成の結合組成物との間の相互作用を指す。この相互作用は、タンパク質の特定の構造（例えば抗原決定基すなわちエピトープ）であって結合分子が認識する構造の有無に依存する。例えば、或る抗体がエピトープ「A」に対して特異的である場合、結合していない標識した「A」と抗体とを含む或る反応の中に、エピトープAを持つポリペプチドが、あるいは結合していない無標識の「A」が存在すると、抗体と結合する標識Aの量が減少する。

【0222】

用語「実質的に精製された」は、自然の環境から取り除かれてから、単離あるいは分離された核酸配列あるいはアミノ酸配列であって、自然に結合している組成物が少なくとも約60%除去されたものであり、好ましくは約75%以上除去、最も好ましくは90%以上除去されたものを指す。 20

【0223】

「置換」とは、1つ以上のアミノ酸残基またはヌクレオチドを、それぞれ別のアミノ酸残基またはヌクレオチドに置き換えることである。

【0224】

用語「基板」は、任意の好適な固体あるいは半固体の支持物を指し、膜およびフィルタ、チップ、スライド、ウェハ、ファイバー、磁気または非磁気ビーズ、ゲル、チューブ、プレート、ポリマー、微小粒子、毛細管が含まれる。基板は、ウェル、溝、ピン、チャンネル、孔など、様々な表面形態を有することができ、基板表面にはポリヌクレオチドやポリペプチドが結合する。 30

【0225】

「転写イメージ(transcript image)」または「発現プロファイル」は、所定条件下での所定時間における、或る特定の細胞タイプまたは組織による、遺伝子発現の集合的パターンを指す。

【0226】

「形質転換(transformation)」とは、外因性DNAが、或る受容細胞に導入されるプロセスを言う。形質転換は、本技術分野で知られている種々の方法に従って自然条件または人工条件下で生じ得るものであり、外来性の核酸配列を原核宿主細胞または真核宿主細胞に挿入する、任意の既知の方法を基にし得る。形質転換の方法は、形質転換する宿主細胞の種類によって選択する。限定するものではないが形質転換方法には、ウイルス感染、電気穿孔法（エレクトロポレーション）、熱ショック、リポフェクションおよび微粒子銃を用いる方法がある。「形質転換された細胞」には、導入されたDNAが、自律的に複製するプラスミドとしてあるいは宿主染色体の一部として複製可能である、安定的に形質転換された細胞が含まれる。更に、限られた期間に一過的に導入DNA若しくは導入RNAを発現する細胞も含まれる。 40

【0227】

ここで用いる「遺伝形質転換体(transgenic organism)」とは任意の生物体であり、限定 50

するものではないが動植物を含み、生物体の1個以上の細胞が、ヒトの介入によって、例えば本技術分野で公知のトランスジェニック技術によって導入された異種核酸を有するものである。細胞への核酸の導入は、直接または間接的に、細胞の前駆物質に導入することによって行う。これは、計画的な遺伝子操作によって、例えば微量注射法によってあるいは組換えウイルスの導入によって行う。或いは核酸の導入は、組換えウイルスベクター、例えばレンチウイルスベクターを感染させて成し得る (Lois, C. 他(2002) Science 295: 868-872)。遺伝子操作の語は、古典的な交雑育種あるいはin vitro受精を指すものではなく、組換えDNA分子の導入を指す。本発明に基づいて予期される遺伝形質転換体には、バクテリア、シアノバクテリア、真菌および動植物がある。本発明の単離されたDNAは、本技術分野で知られている方法、例えば感染、形質移入、形質転換またはトランス接合によって宿主に導入することができる。本発明のDNAをこのような生物体に移入する技術はよく知られており、前出のSambrook 他(1989)などの参考文献に記載されている。

【0228】

特定の核酸配列の「変異体」とは、核酸配列1本の或る長さ全体について、該特定核酸配列に対し少なくとも40%の配列同一性を有する核酸配列として決定された配列である。決定には、デフォルトパラメータに設定した「BLAST 2 Sequences」ツールVersion 2.0.9 (1999年5月7日)を用いてblastnを実行する。このような核酸対は、所定の長さに対して、例えば少なくとも50%、60%、70%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%またはそれ以上の配列同一性を示し得る。変異体は、例えば、「対立遺伝子」変異体(上述)または「スプライス」変異体、「種」変異体、「多型」変異体と記述され得る。スプライス変異体は参照分子との顕著な同一性を有し得るが、mRNAプロセッシング中のエキソン群の選択的スプライシングによって通常、より多数またはより少数のヌクレオチドを有することになる。対応するポリペプチドは、追加機能ドメイン群を有するか、あるいは参照分子には存在するドメイン群が欠落していることがある。種変異体は、種によって異なるポリヌクレオチド配列である。結果的に生じるポリペプチドは通常、相互に有意のアミノ酸同一性を持つ。多型性変異体は、或る種の個体間における、或る特定遺伝子のポリヌクレオチド配列内での変異である。多型変異体にはまた、ポリヌクレオチド配列の1つのヌクレオチド塩基が異なる「1塩基多型性」(SNP)も含み得る。SNPの存在は、例えば特定の個体群、病状または病状性向を示し得る。

【0229】

特定のポリペプチド配列の「変異体」とは、ポリペプチド配列1本の或る長さ全体について、該特定ポリペプチド配列に対し少なくとも40%の配列同一性を有するポリペプチド配列として決定された配列である。決定には、デフォルトパラメータに設定した「BLAST 2 Sequences」ツールVersion 2.0.9 (1999年5月7日)を用いてblastpを実行する。このようなポリペプチド対は、そのポリペプチドの一方の所定の長さに対して、例えば少なくとも50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%またはそれ以上の配列同一性を示し得る。

【0230】

(発明)

本発明は、新規のヒト細胞成長、分化および細胞死関連タンパク質(CGDD)群と、CGDDをコードするポリヌクレオチド群との発見および、これらの組成を、癌など細胞増殖異常、発達障害、神経疾患、自己免疫/炎症疾患、生殖障害、及び胎盤障害の、診断、治療並びに予防に利用する方法の発見に基づく。

【0231】

表1は、本発明の完全長ポリヌクレオチド配列およびポリペプチド配列の命名の概略である。各ポリヌクレオチドおよびその対応するポリペプチドは、1つのIncyteプロジェクト識別番号(IncyteプロジェクトID)に相関する。各ポリペプチド配列は、ポリペプチド配列識別番号(ポリペプチドSEQ ID NO:)とIncyteポリペプチド配列番号(IncyteポリペプチドID)によって表示した。各ポリヌクレオチド配列は、ポリヌクレオチド配列識別番号

(ポリヌクレオチドSEQ ID NO:)とIncyteポリヌクレオチドコンセンサス配列番号 (IncyteポリヌクレオチドID) によって表示した。

【0232】

表2は、GenBankタンパク質 (genpept) データベースに対するBLAST分析で同定した、本発明のポリペプチド群に相同な配列群を示す。列1および列2はそれぞれ、本発明の各ポリペプチドに対するポリペプチド配列識別番号 (ポリペプチド SEQ ID NO:)と、それに対応するIncyteポリペプチド配列番号 (IncyteポリペプチドID)を示す。列3は、最も近いGenBank相同体のGenBank識別番号 (GenBank ID NO:)を示す。列4は、各ポリペプチドとその相同体1つ以上との間の一致に関する確率スコアを示す。列5は、該当箇所には適当な引用を示すと共にGenBank相同体1つ以上の注釈 (annotation)を示し、これらは全て引用を以て本明細書の一部とする。 10

【0233】

表3は、本発明のポリペプチドの多様な構造的特徴を示す。列1および列2はそれぞれ、本発明の各ポリペプチドに対するポリペプチド配列識別番号 (SEQ ID NO:)と、それに対応するIncyteポリペプチド配列番号 (IncyteポリペプチドID)を示す。列3は、各ポリペプチドのアミノ酸残基数を示す。列4および列5はそれぞれ、GCG配列分析ソフトウェアパッケージのMOTIFSプログラム (Genetics Computer Group, Madison WI)によって決定された、リン酸化およびグリコシル化の可能性のある部位を示す。列6は、シグネチャ配列、ドメイン、およびモチーフを含むアミノ酸残基を示す。列7は、タンパク質の構造/機能の分析のための分析方法を示し、該当箇所には更に、分析方法に利用した検索可能なデータベースを示す。 20

【0234】

表2および3は共に、本発明の各々のポリペプチドの特性を要約しており、それら特性が、請求の範囲に記載されたポリペプチドが細胞成長、分化および細胞死関連タンパク質であることを確立している。

【0235】

例えばSEQ ID NO:1は、残基M1から残基L1738までが、マウスのユビキチン蛋白リガーゼE3 α (GenBank ID g3170887)に対して92%同一であると、Basic Local Alignment Search Tool (BLAST)で判定された (表2参照)。BLAST確率スコアは0.0であり、これは観測されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO:1はまた、1つの推定ZnフィンガーをN-recogninドメインに有するが、これは、隠れマルコフモデル (HMM)を基にした保存されたタンパク質ファミリードメインのPFAMデータベースにおいて、統計的に有意な一致を検索して決定された (表3参照)。更なるBLAST解析からのデータは、SEQ ID NO:1がユビキチン蛋白質リガーゼであることを示す更に確証的な証拠を提供する。 30

【0236】

別の例でSEQ ID NO:3は、残基M1から残基D854までが、マウスのユビキチン蛋白リガーゼNedd4-2 (GenBank ID g12656270)に対して88%同一であると、Basic Local Alignment Search Tool (BLAST)で判定された (表2参照)。BLAST確率スコアは0.0であり、これは観測されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO:3はまた、HECT (ユビキチン転移酵素)ドメインとWWドメインを有することが、隠れマルコフモデル (HMM)を基にした保存されたタンパク質ファミリードメインのPFAMデータベースにおいて、統計的に有意な一致を検索して決定された (表3参照)。BLIMPSおよびMOTIFS解析からのデータは、SEQ ID NO:3がユビキチン蛋白質リガーゼであることを示す更に確証的な証拠を提供する。 40

【0237】

別の例でSEQ ID NO:5は、残基M27から残基D538までが、ヒトのシスプラチン耐性関連タンパク質CRR9p (GenBank ID g12248402)に対して100%同一であると、Basic Local Alignment Search Tool (BLAST)で判定された (表2参照)。BLAST確率スコアは $8.4e-281$ であり、これは観測されたポリペプチド配列を偶然に得る確率を示す。更なるBLAST解析からのデー 50

タは、SEQ ID NO:5がアポトーシス関連タンパク質であることを示す更に確証的な証拠を提供する。

【0238】

別の例においてSEQ ID NO:9は、残基M1から残基S710までが、マウスのRalGDS様タンパク質3(GenBank ID g8650435)に対して80%同一であると、Basic Local Alignment Search Tool (BLAST)で判定された(表2参照)。BLAST確率スコアは $1.5e-299$ であり、これは観測されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO:9はまた、1つのRas会合(RalGDS/AF-6)ドメインと1つのRasGEFドメインを有するが、これは、隠れマルコフモデル(HMM)を基にした保存されたタンパク質ファミリードメインのPFAMデータベースにおいて、統計的に有意な一致を検索して決定された(表3参照)。更なるBLAST解析にPRODOMおよびDOMOデータベースを用いてのデータは、SEQ ID NO:9がグアニンスクレオチド解離因子である、更に実証的な証拠を提供する。

【0239】

別の例においてSEQ ID NO:12は、残基A64から残基Y365までが77%、残基M1から残基D109までが100%、Sgt1 (GenBank ID g4809026)に対して同一であると、Basic Local Alignment Search Tool (BLAST)で判定された(表2参照)。BLAST確率スコアは $3.2e-121$ であり、これは観測されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO:12はまた、テトラトリコペプチド(tetratricopeptide) (TPR)ドメイン群を残基A45~N78と残基S79~T112に有することが、隠れマルコフモデル(HMM)を基にした保存されたタンパク質ファミリードメインのPFAMデータベースにおいて、統計的に有意な一致を検索して決定された(表3参照)。TPRリピーターは、蛋白間相互作用を媒介すると思われる。また、有糸分裂に関わる多数の蛋白に見られる。更に、SPSCANの同定では、潜在的シグナルペプチドが残基M1~A68に在る。

【0240】

別の例においてSEQ ID NO:13は、残基M1から残基K365までが、ヒトのプロトオンコジーンWnt-5A(GenBank ID g348918)に対して100%同一であると、Basic Local Alignment Search Tool (BLAST)で判定された(表2参照)。BLAST確率スコアは $2.5e-205$ であり、これは観測されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO:13はまた、1つのwnt-1ファミリードメインを有するが、これは、隠れマルコフモデル(HMM)を基にした保存されたタンパク質ファミリードメイン群のPFAMデータベースにおいて、統計的に有意な一致を検索して決定された(表3参照)。BLIMPS解析、MOTIFS解析、およびPROFILESCAN解析よりのデータは、SEQ ID NO:13がwnt-1ファミリー蛋白質(wntファミリーの分泌性糖タンパクのメンバー)である、更に実証的な証拠を提供する。

【0241】

別の例においてSEQ ID NO:16は、残基A18から残基F1014までが、ヒトのサイクリンE結合タンパク質1(GenBank ID g6630609)に対して50%同一であると、Basic Local Alignment Search Tool (BLAST)で判定された(表2参照)。BLAST確率スコアは $3.6e-252$ であり、これは観測されたポリペプチド配列を偶然に得る確率を示す。SEQ ID NO:16はまた、1つのHECT(ユビキチン転移酵素)ドメインと、1つの染色体凝縮調節因子(RCC1)蛋白ドメインを有し、これは、隠れマルコフモデル(HMM)を基にした、保存されたタンパク質ファミリードメイン群のPFAMデータベースにおいて、統計的に有意な一致を検索して判定された(表3参照)。BLIMPS、MOTIFS、及びPROFILESCAN解析よりのデータは、SEQ ID NO:16がサイクリン結合蛋白質である、さらに実証的な証拠を提供する。

【0242】

別の例においてSEQ ID NO:17は、残基M1から残基S1462までが、ヒトのシクロフィリン関連蛋白(GenBank ID g5923891)に対して99%同一であると、Basic Local Alignment Search Tool (BLAST)で判定された(表2参照)。BLAST確率スコアは 0.0 であり、これは観測されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO:17はまた、1つのシクロフィリン型ペプチジルプロリルシストランス異性化酵素ドメインを有するが、これは、隠れマルコフモデル(HMM)を基にした保存されたタンパク質ファミリ

ードメインのPFAMデータベースにおいて、統計的に有意な一致を検索して決定された（表3参照）。BLIMPS、MOTIFS、及びPROFILESCAN解析よりのデータは、SEQ ID NO:17がシクロフィリン関連蛋白質である、さらに実証的な証拠を提供する。

【0243】

別の例においてSEQ ID NO:19は、残基K3から残基S175までが34%、残基R40から残基Q327まで26%、ヒトのアポトーシス性プロテアーゼ活性化因子1 (GenBank ID g2330015)に対して同一であると、Basic Local Alignment Search Tool (BLAST)で判定された(表2参照)。BLAST確率スコアは $8.3e-21$ であり、これは観測されたポリペプチド配列を偶然に得る確率を示す。SEQ ID NO:19はまた、1つのSAMドメインと、G蛋白 β WD-40リピート群を有するが、これは、隠れマルコフモデル (HMM) を基にした保存されたタンパク質ファミリードメイン群のPFAMデータベースにおいて、統計的に有意な一致を検索して決定された（表3参照）。BLIMPS、MOTIFS、PROFILESCAN解析のデータが提供する更に実証的な証拠によれば、SEQ ID NO:19は、多重 β G蛋白WD-40シグネチャ (Apaf-1に類似)を持つ。SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6-8, SEQ ID NO:10-11, SEQ ID NO:14-15, SEQ ID NO:18, SEQ ID NO:20-21の分析と注釈も同様に行った。SEQ ID NO:1-21の解析用のアルゴリズム及びパラメータを表7に記載した。

10

【0244】

表4に示すように、本発明の完全長ポリヌクレオチド配列は、cDNA配列またはゲノムDNA由来のコード（エキソン）配列を用いて、あるいはこれら2種類の配列を任意に組合せて構築した。列1は本発明の各ポリヌクレオチドに対するポリヌクレオチド配列識別番号（ポリヌクレオチドSEQ ID NO）および対応するIncyteポリヌクレオチドコンセンサス配列番号（Incyte ID）、および塩基対の各ポリヌクレオチド配列の長さを示している。列2は、本発明の完全長ポリヌクレオチド配列を構築するのに用いたcDNA配列および／またはゲノム配列の、また、例えばSEQ ID NO:22-42を同定するため、或いはSEQ ID NO:22-42と、関連するポリヌクレオチド配列群とを区別するためのハイブリダイゼーション技術または増幅技術に有用なポリヌクレオチド配列の断片の、開始ヌクレオチド（5'）位置および終了ヌクレオチド（3'）位置を示す。

20

【0245】

表4の列2に記載したポリヌクレオチド断片は、特に例えば組織特異的cDNAライブラリ群やプールしたcDNAライブラリ群に由来するIncyte cDNA群を指す場合もある。或いは列2に記載したポリヌクレオチド断片は、完全長ポリヌクレオチド配列の構築（アセンブリ）に寄与したGenBank cDNAまたはESTを指す場合もある。更に、列2に記載したポリヌクレオチド断片は、ENSEMBL (The Sanger Centre、英国ケンブリッジ) データベースに由来する配列を同定する場合もある（すなわち「ENST」の命名を含む配列）。あるいは、列2に記載したポリヌクレオチド断片は、NCBI RefSeq Nucleotide Sequence Records データベースに由来する場合もあり（すなわち「NM」または「NT」の命名を含む配列）、またNCBI RefSeq Protein Sequence Recordsに由来する場合もある（すなわち「NP」の命名を含む配列）。または列2に記したポリヌクレオチド断片は、「エキソンスティッチング (exon-stitching)」アルゴリズムにより結び合わせたcDNAおよびGenscan予測エキソン群の両方からなる構築物を指す場合がある。例えば、ポリヌクレオチド配列の内、FL_XXXXXX_N₁_N₂_YYYYY_N₃_N₄として同定される配列は「スティッチされた」配列であり、その内、XX XXXXは該アルゴリズムが適用される配列のクラスターの識別番号であり、YYYYYは該アルゴリズムが作成する予測の数であり、N₁, N₂, N₃, ...がある場合には、解析中に手動で編集された特定のエキソン群を表す（実施例5参照）。または、列2のポリヌクレオチド断片は「エキソンストレッチング (exon-stretching)」アルゴリズムにより結び合わせたエキソンの構築物を指す場合もある。例えば、ポリヌクレオチド配列の内、FLXXXXXX_gAAAAA_gBBBBB_1_Nとして同定される配列は「ストレッチ」配列の識別番号である。ここでXXXXXXはIncyteプロジェクト識別番号、gAAAAAは「エキソンストレッチング」アルゴリズムを適用したヒトゲノム配列のGenBank識別番号、gBBBBBは一番近いGenBankタンパク質相同体のGenBank識別番号、またはNCBI RefSeq識別番号、Nは特定のエキソンを指す（実施例5を

30

40

50

参照)。或るRefSeq配列が「エキソントレッチング」アルゴリズムのためのタンパク質相同体として使用された場合では、「NM」、「NP」、または「NT」によって表されるRefSeq識別子が、GenBank識別子（すなわち、gBBBBB）の代わりに使用され得る。

【0246】

あるいは、接頭コードは、手動で編集された構成配列、ゲノムDNA配列から予測された構成配列、または組み合わされた配列解析方法に由来する構成配列を同定する。次の表は、構成配列の接頭コードと、接頭コードに対応する配列分析方法の例を列記する（実施例4と5を参照）。

接頭コード	解析タイプやプログラムの例
GNN, GFG, ENST	例えば、GENSCAN (Stanford University, CA, USA) または FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK) を用いた、ゲノム配列群からのエキソン予測
GBI	手動で編集された、ゲノム配列群の解析
FL	スティッチまたはストレッチゲノム配列（ <u>実施例5</u> 参照）
INCY	ゲノムへのEST配列群のマッピングからの、完全長転写物とエキソンとの予測。エキソン群と生じる転写物とを予測するために、ゲノム位置とEST構成のデータが組み合わされる。

10

20

【0247】

場合によっては、最終コンセンサスポリヌクレオチド配列を確認するために、表4に示すような配列カバレッジと重複するIncyte cDNAカバレッジが得られたが、該当するIncyte cDNA識別番号は示さなかった。

30

【0248】

表5は、Incyte cDNA配列を用いて構築された完全長ポリヌクレオチド配列のための代表的なcDNAライブラリを示している。代表的なcDNAライブラリとはIncyte cDNAライブラリであり、これは、最も頻繁にはIncyte cDNA配列群によって代表されるが、これらは、上記のポリヌクレオチド配列を構築および確認するために用いられた。cDNAライブラリを作製するために用いた組織およびベクターを表5に示し、表6で説明している。

【0249】

本発明にはCGDD変異体も含まれる。好適なCGDD変異体は、CGDDの機能的或いは構造的特徴の少なくとも1つを有し、かつ該CGDDアミノ酸配列に対して少なくとも約80%のアミノ酸配列同一性、或いは少なくとも約90%のアミノ酸配列同一性、更には少なくとも約95%のアミノ酸配列同一性を有する配列である。

40

【0250】

本発明はCGDDをコードするポリヌクレオチド群をも含む。特定の実施態様において、本発明は、CGDDをコードする、SEQ ID NO:22-42からなる一群から選択された1配列を持つポリヌクレオチド配列を提供する。SEQ ID NO:22-42のポリヌクレオチド配列は、配列表に示されているように等価RNA配列をも含むが、そこでは窒素塩基チミンの出現はウラシルに置換され、糖のバックボーンはデオキシリボースではなくてリボースから構成されている。

【0251】

50

CGDDをコードするポリヌクレオチド配列の変異体も、本発明に含まれる。詳細には、このような変異体ポリヌクレオチド配列は、CGDDをコードするポリヌクレオチド配列との、少なくとも約70%のポリヌクレオチド配列同一性、あるいは少なくとも約85%のポリヌクレオチド配列同一性、更には少なくとも約95%のポリヌクレオチド配列同一性を持つこととなる。本発明の或る態様では、SEQ ID NO:22-42からなる群から選択されたアミノ酸配列と少なくとも約70%、或いは少なくとも約85%、または少なくとも約95%の一致率を有するようなSEQ ID NO:22-42からなる群から選択された配列を有するポリヌクレオチド配列の変異体を含む。上記の任意のポリヌクレオチドの変異体は、CGDDの機能的若しくは構造的特徴を少なくとも1つ有するアミノ酸配列をコードし得る。

【0252】

更に又は別の例で本発明のポリヌクレオチド変異体は、CGDDをコードするポリヌクレオチド配列のスプライス変異体である。或るスプライス変異体は、CGDDをコードするポリヌクレオチド配列との顕著な配列同一性を持つ部分複数を有し得るが、mRNAプロセッシング中のエキソン群の選択的スプライシングによって生ずる、配列の数ブロックの付加または欠失により、通常、より多数またはより少数のヌクレオチドを有することになる。或るスプライス変異体には、約70%未満、または約60%未満、あるいは約50%未満のポリヌクレオチド配列同一性が、CGDDをコードするポリヌクレオチド配列との間で全長に渡って見られるが、このスプライス変異体のいくつかの部分には、CGDDをコードするポリヌクレオチド配列の各部との、少なくとも約70%、あるいは少なくとも約85%、または少なくとも約95%、なおまたは100%の、ポリヌクレオチド配列同一性を有することとなる。例えばSEQ ID NO:42は、SEQ ID NO:41の配列からなるポリヌクレオチドのスプライス変異体である。上記したスプライス変異体は何れも、CGDDの機能的或いは構造的特徴の少なくとも1つを有する或るアミノ酸配列をコードし得る。

【0253】

CGDDをコードする種々のポリヌクレオチド配列が遺伝暗号の縮重によって作り出され、中には、既知のいかなる天然遺伝子のポリヌクレオチド配列群とも最小の類似性しか有しない配列もあることは、当業者には理解されよう。したがって本発明には、可能コドン選択に基づく組合せの選択によって産出し得るあらゆる可能なポリヌクレオチド配列のバリエーションを網羅し得る。天然CGDDのポリヌクレオチド配列に適用される標準的なトリプレット遺伝暗号を基に、これらの組合せが成される。このような全ての変異は明確に開示されていると考慮されたい。

【0254】

CGDDとその変異体とをコードするヌクレオチド配列は一般に、好適に選択されたストリンジェンシー条件下で天然CGDDのヌクレオチド配列とハイブリダイズ可能であるが、非天然コドン群を含めるなどの実質的に異なるコドン使用を有する、CGDD或いはその誘導体をコードするヌクレオチド配列を作り出すことは、有益であり得る。宿主が特定コドンを利用する頻度に基づいて、特定の真核宿主または原核宿主に発生するペプチドの発現率を高めるようにコドンを選択し得る。コードされるアミノ酸配列を改変せずに、CGDDおよびその誘導体をコードするヌクレオチド配列を実質的に改変する別の理由には、天然配列から作られる転写物より例えば長い半減期など好ましい特性を備えるRNA転写物を作ることもある。

【0255】

CGDDとその誘導体とをコードするDNA配列またはそれらの断片の、完全に合成化学による作製も、本発明に含まれる。作製後にこの合成配列を、当分野で周知の試薬類を用いて、種々の入手可能な発現ベクター及び細胞系の何れの中にも挿入できる。更に、合成化学を用いて、CGDDをコード、またはその任意の断片をコードする或る配列の中に突然変異を導入できる。

【0256】

更に本発明には、種々のストリンジェンシー条件下で、請求項に記載されたポリヌクレオチド配列、特に、SEQ ID NO:22-42及びそれらの断片群にハイブリダイズ可能なポリヌク

10

20

30

40

50

レオチド配列群が含まれる（例えば、Wahl, G.M.およびS.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511を参照）。アニーリングおよび洗浄条件を含むハイブリダイゼーションの条件は、「定義」に記載した。

【0257】

DNAシーケンシングの方法は当分野では公知であり、本発明のいずれの実施例も、DNAシーケンシング方法を用い得る。DNAシーケンシング方法には酵素を用い得る。例えばDNAポリメラーゼ1のクレノウ断片、SEQUENASE (US Biochemical, Cleveland OH)、Taqポリメラーゼ (Applied Biosystems)、熱安定性T7ポリメラーゼ (Amersham, Pharmacia Biotech, Piscataway NJ) を用い得る。あるいは、例えばELONGASE増幅システム (Life Technologies, Gaithersburg MD) において見られるように、ポリメラーゼと校正エキソヌクレアーゼとを併用し得る。好適には、MICROLAB2200液体転移システム (Hamilton, Reno, NV)、PTC200サーマルサイクラー (MJ Research, Watertown MA) およびABI CATALYST 800サーマルサイクラー (Applied Biosystems) などの装置を用いて配列の準備を自動化する。次に、ABI 373あるいは377 DNAシーケンシングシステム (Applied Biosystems)、MEGABACE 1000 DNAシーケンシングシステム (Molecular Dynamics, Sunnyvale CA) または当分野で公知の他のシステムを用いてシーケンシングを行う。結果として得た配列を、当分野で周知の種々のアルゴリズムを用いて分析する（例えば、Ausubel, F.M. (1997) *Short Protocols in Molecular Biology*, John Wiley & Sons, New York NY, 7.7ユニット、Meyers, R.A. (1995) *Molecular Biology and Biotechnology*, Wiley VCH, New York NY, 856-853ページを参照）。

10

20

【0258】

CGDDをコードする核酸配列を伸長し、プロモーターや調節エレメントなど、上流にある配列を検出するには、当分野で周知の、PCR法をベースにした種々の方法と、部分的ヌクレオチド配列とを用い得る。例えば、使用し得る方法の1つである制限部位PCR法は、ユニバーサルプライマーおよびネステッドプライマーを用いてクローニングベクター内のゲノムDNAからの未知の配列を増幅する方法である（例えばSarkar, G. (1993) *PCR Methods Appl.* 2:318-322を参照）。別の方法に逆PCR法があり、これは多岐の方向に伸長するプライマー群を用いて、環状化した鋳型から未知の配列を増幅する方法である。鋳型は、或る既知のゲノム遺伝子座およびその周辺の配列群からなる制限酵素断片群から得る（例えばTriglia, T. 他(1988) *Nucleic Acids Res.* 16:8186を参照）。第3の方法としてキャプチャーPCR法があり、これにはヒトおよび酵母人工染色体DNAの既知の配列群に隣接するDNA断片群をPCR増幅する方法を含む（例えばLagerstrom, M. 他(1991) *PCR Methods Appl.* 1:111-119を参照）。この方法では、PCRを行う前に、複数の制限酵素の消化およびライゲーション反応を用い、未知の配列領域内に組換え二本鎖配列を挿入し得る。未知の配列群を検索するために用い得る、他の複数の方法も当分野で既知である（例えばParker, J.D. 他(1991) *Nucleic Acids Res.* 19:3055-3060を参照）。更に、PCR、ネステッドプライマー類、およびPROMOTERFINDERライブラリ類 (Clontech, Palo Alto CA) を用いて、ゲノムDNAをウォーキングし得る。この手順は、ライブラリ類をスクリーニングする必要がなく、イントロン/エキソン接合部の発見に有用である。全てのPCRベースの方法では、市販ソフトウェア、例えばOLIGO 4.06プライマー分析ソフトウェア (National Biosciences, Plymouth MN) 或いは別の好適なプログラムを用いて、長さが約22~30ヌクレオチド、GC含有率が約50%以上で、温度約68℃~72℃で鋳型に対してアニーリングするように、プライマー群を設計し得る。

30

40

【0259】

完全長cDNA群をスクリーニングする際は、より大きなcDNA群を含むようにサイズ選択されたライブラリ群を用いるのが好ましい。更に、ランダムプライマーのライブラリ群は、しばしば遺伝子群の5'領域を有する配列を含み、オリゴd(T)ライブラリが完全長cDNAを作製できない状況に対して好適である。ゲノムライブラリ群は、5'非転写調節領域への、配列の伸長に有用であろう。

【0260】

50

市販のキャピラリー電気泳動システム群を用いて、シークエンシングまたはPCR産物のサイズを分析し、またはそのヌクレオチド配列を確認し得る。具体的には、キャピラリーシークエンシングは、電気泳動による分離のための流動性ポリマーと、4つの異なるヌクレオチドに特異的であるような、レーザーで活性化される蛍光色素と、発光された波長の検出に利用するCCDカメラとを有し得る。出力／光の強度は、適切なソフトウェア（Applied Biosystems社のGENOTYPER、SEQUENCE NAVIGATORなど）を用いて電気信号に変換し得る。サンプルのロードからコンピュータ分析および電子データ表示までの全プロセスがコンピュータ制御可能である。キャピラリー電気泳動法は、特定のサンプルに少量しか存在しないようなDNA小断片のシークエンシングに特に適している。

【0261】

CGDDをコードするポリヌクレオチド配列またはその断片を、本発明の別の実施態様では、CGDD、その断片または機能的等価物を適切な宿主細胞内に発現させる組換えDNA分子にクローニングし得る。CGDDの発現に、遺伝暗号固有の縮重により産生されうる、実質的に同じあるいは機能的に等価のアミノ酸配列をコードする別のDNA配列群を利用可能である。

【0262】

CGDDをコードする配列群を種々の目的で改変するために、当分野で一般的に既知の複数の方法を用いて、本発明のヌクレオチド配列群を組換えることができる。この組換えの多様な目的には、遺伝子産物のクローン化の、あるいはプロセッシングおよび／または発現のモディフィケーションが含まれるが、これらに限定されるものではない。遺伝子断片と合成オリゴヌクレオチドとの、ランダムなフラグメンテーションおよびPCR再アセンブリによるDNAシャッフリングを用い、ヌクレオチド配列を組換え得る。例えば、オリゴヌクレオチドを介した部位特異的変異誘導を利用して、新規な制限部位の作製、グリコシル化パターンの改変、コドン優先の変更、スプライス変異体の生成などを起こす突然変異を導入し得る。

【0263】

本発明のヌクレオチドは、MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; 米国特許第5,837,458号; Chang, C.-C. 他(1999) Nat. Biotechnol. 17:793-797; Christians, F.C. 他(1999) Nat. Biotechnol. 17:259-264; Cramer, A. 他(1996) Nat. Biotechnol. 14:315-319)などのDNAシャッフリング技術の対象となり得る。これにより、生物活性または酵素活性、あるいは他の分子や化合物と結合する能力など、CGDDの生物学的特性を改変あるいは改良し得る。DNAシャッフリングは、PCRを介する遺伝子断片組換えを用いて遺伝子変異体のライブラリを作製するプロセスである。ライブラリはその後、所望の特性を持つ遺伝子変異体群を同定する、選択またはスクリーニングの手順を経る。続いて、これら好適な変異体をプールし、更に反復してDNAシャッフリングおよび選択／スクリーニングを行い得る。かくして、「人工的な」育種および急速な分子の進化によって、多様な遺伝子が作られる。例えば、ランダムポイント突然変異を持つ単一の遺伝子の断片を組換えて、スクリーニングし、その後所望の特性が最適化されるまでシャッフリングし得る。あるいは、所定の遺伝子の断片群を同種または異種のいずれかから得た同一遺伝子ファミリーの相同遺伝子の断片と組換え、それによって天然に存在する複数の遺伝子の遺伝多様性を、方向付けした制御可能な方法で最大化させることができる。

【0264】

CGDDをコードする配列は、別の実施態様では、当分野で周知の化学的方法を用いて、全体或いは一部を合成可能である（例えば、Caruthers, M.H. 他(1980) Nucleic Acids Symp. Ser. 7:215-223; および Horn, T. 他(1980) Nucleic Acids Symp. Ser. 7:225-232を参照）。CGDD自体またはその断片の合成には、或いは化学的方法を用い得る。例えば、種々の液相または固相技術を用いてペプチド合成を行い得る（例えばCreighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, 55-60ページ; および Roberge, J.Y. 他(1995) Science 269:202-204を参照）。自動合成はABI 431A ペプチドシンセサイザ（Applied Biosystems）を用いて達成し得る。CGDDのアミノ酸配列、または任意のその一部は更に、直接合成の際に改変することにより、及び／または他の

10

20

30

40

50

タンパク質からの配列群または任意のその一部と組合せることにより、変異体ポリペプチドを作製、または、或る天然ポリペプチドの1配列を有するポリペプチドを作製し得る。

【0265】

このペプチドは、分取用高速液体クロマトグラフィーを用いて実質的に精製し得る（例えばChiez, R.M.およびF.Z. Regnier (1990) *Methods Enzymol.* 182:392-421を参照）。この合成ペプチドの組成は、アミノ酸分析またはシーケンシングによって確認できる（例えば前出のCreighton, 28-53ページを参照）。

【0266】

生物活性CGDDの発現には、CGDDをコードするヌクレオチド配列またはその誘導体を好適な発現ベクターに挿入し得る。好適な発現ベクターとは、好適な宿主に挿入されたコーディング配列の転写および翻訳の制御に必要なエレメント群を持つベクターである。CGDDをコードするポリヌクレオチド配列、およびベクターにおける調節配列（例えばエンハンサー、構成型および発現誘導型のプロモーター、5'および3'の非翻訳領域など）が、必要なエレメントに含まれる。必要なエレメント群は、強度および特異性が様々である。CGDDをコードする配列群の、より効率的な翻訳を、特異的な開始シグナル類を用いて達成することもできる。開始シグナルの例には、ATG開始コドンと、コザック配列など近傍の配列とが含まれる。CGDDをコードする配列群、その開始コドン、および上流の調節配列群が、好適な発現ベクターに挿入された場合は、更なる転写制御シグナルや翻訳制御シグナルは必要ないこともある。しかしながら、コード配列あるいはその断片のみが挿入された場合は、インフレームATG開始コドンなど外来性の翻訳制御シグナルが発現ベクターに含まれるようにすべきである。外因性の翻訳エレメント群および開始コドン群は、様々な天然物および合成物を起源とし得る。用いられる特定の宿主細胞系に好適なエンハンサーを含めることで発現の効率を高めることが可能である（例えばScharf, D. 他(1994) *Results Probl. Cell Differ.* 20:125-162を参照）。

【0267】

CGDDをコードする配列と、好適な転写および翻訳制御エレメントとを持つ発現ベクターを、当業者に周知の方法を用いて作製し得る。これらの方法としては、in vitro組換えDNA技術、合成技術、およびin vivo遺伝子組換え技術がある（例えばSambrook, J.他(1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview NY, 4, 8, および 16-17章; Ausubel, F.M.他(1995) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York NY, 9, 13, および 16章を参照）。

【0268】

CGDDをコードする配列を、種々の発現ベクター／宿主系を利用して保持および発現し得る。限定するものではないがこのような発現ベクター／宿主系には、組換えバクテリオファージ、プラスミドまたはコスミドDNA発現ベクターで形質転換させた細菌や、酵母菌発現ベクターで形質転換させた酵母菌など微生物や、ウイルス発現ベクター（例えばバキュロウイルス）に感染した昆虫細胞系や、ウイルス発現ベクター（例えばカリフラワーモザイクウイルス、CaMVまたはタバコモザイクウイルス、TMV）または細菌発現ベクター（例えばTiプラスミドまたはpBR322プラスミド）で形質転換させた植物細胞系、動物細胞系がある（例えば前出Sambrook; 前出Ausubel; Van Heeke, G. および S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. 他(1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V.他(1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; 『マグローヒル科学技術年鑑』(*The McGraw Hill Yearbook of Science and Technology*) (1992) McGraw Hill, New York NY, 191-196ページ; Logan, J. および T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; および Harrington, J.J. 他(1997) *Nat. Genet.* 15:345-355を参照)。レトロウイルス、アデノウイルス、ヘルペスウイルスまたはワクシニアウイルス由来の発現ベクター、または種々の細菌性プラスミド由来の発現ベクターを用いて、ヌクレオチド配列を標的器官、組織または細胞集団へ送達することができる（例えばDi Nicola, M.他(1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M.他(1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M

.他 (1985) Nature 317(6040):813-815; McGregor, D.P.他 (1994) Mol. Immunol. 31(3):219-226; 並びに Verma, I.M. および N. Somia (1997) Nature 389:239-242を参照)。本発明は使用される宿主細胞によって限定されるものではない。

【0269】

CGDDをコードするポリヌクレオチド配列の使用目的に応じて多数のクローニングベクターおよび発現ベクターを、細菌系で選択し得る。CGDDをコードするポリヌクレオチド配列の慣例的なクローニング、サブクローニング、増殖には例えば、PBLUESCRIPT (Stratagene, La Jolla CA) またはPSPORT1 プラスミド (Life Technologies) などの多機能の大腸菌ベクターを用いることができる。CGDDをコードする配列をベクターのマルチクローニング部位にライゲーションするとlacZ 遺伝子が破壊され、組換え分子を含む形質転換された細菌の同定のための比色スクリーニング法が可能となる。更にこれらのベクターは、クローニングされた配列における *in vitro* 転写、ジデオキシのシーケンシング、ヘルパーファージによる一本鎖のレスキュー、入れ子欠失の生成にも有用であろう (例えば、Van Heeke, G.およびS.M. Schuster (1989) J. Biol. Chem. 264:5503-5509を参照)。CGDDが多量に、例えば抗体の産生などに必要な場合は、CGDDの発現をハイレベルで誘導するベクターが使用できる。例えば、強力な誘導SP6バクテリオファージプロモーターまたは誘導T7バクテリオファージプロモーターを持つベクターが使用できる。

【0270】

CGDDの産生には酵母の発現系を使用できる。α 因子、アルコールオキシダーゼ、PGHプロモーターなど、構成型あるいは誘導型のプロモーターを持つ多数のベクターが、出芽酵母菌 (*Saccharomyces cerevisiae*) またはピキア酵母 (*Pichia pastoris*) 内で使用可能である。更に、このようなベクターは、発現したタンパク質の、分泌か細胞内での保持かのどちらかを指示するものであり、安定した増殖のために宿主ゲノムの中に外来配列群を組み込むことを可能にする (例えば前出のAusubel, 1995; Bitter, G.A. 他(1987) Methods Enzymol.153:516-544; および Scorer, C.A.他(1994) Bio/Technology 12:181-184を参照)。

【0271】

CGDDの発現には植物系も用い得る。CGDDをコードする配列の転写は、ウイルスプロモーター、例えば単独であるいはTMV由来のオメガリーダー配列と組合せて用いられるようなCamV由来の35Sおよび19Sプロモーターによって促進し得る (Takamatsu, N. (1987) EMBO J 6:307-311)。あるいは、RuBisCOの小サブユニットなどの植物プロモーター、または熱ショックプロモーターを用い得る (例えばCoruzzi, G.他(1984) EMBO J. 3:1671-1680; Broglie, R.他(1984) Science 224:838843; および Winter, J.他(1991) Results Probl.Cell Differ. 17:85-105を参照)。これらの作製物は、直接DNA形質転換により、または病原体を媒介とする形質移入により、植物細胞内に導入し得る (例えば『マグローヒル科学技術年鑑』(The McGraw Hill Yearbook of Science and Technology) (1992) McGraw Hill New York NY,191-196ページを参照)。

【0272】

哺乳類細胞においては、多数のウイルスベースの発現系を利用し得る。アデノウイルスが発現ベクターとして用いられる場合、後期プロモーターと3連リーダー配列とからなるアデノウイルス転写／翻訳複合体に、CGDDをコードする配列群をライゲーションし得る。CGDDを宿主細胞内で発現する感染ウイルスを得るため、アデノウイルスゲノムの非必須E1またはE3領域への挿入を用いる (例えばLogan, J.およびT. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659を参照)。更に、ラウス肉腫ウイルス (RSV) エンハンサーなどの転写エンハンサーを用いて、哺乳動物宿主細胞における発現を増大させ得る。SV40またはEBVをベースにしたベクターを用いてタンパク質を高レベルで発現させることもできる。

【0273】

また、ヒト人工染色体 (HAC) 類を用いて、或るプラスミドに含まれ得る断片やプラスミドから発現し得る断片より大きなDNAの断片群を送達し得る。治療のために約6 kb～10 M

bのHACsが作製され、従来の送達方法（リポソーム、ポリカチオンアミノポリマー、またはベシクル）で送達されている（例えばHarrington, J.J.他(1997) Nat. Genet. 15:345-355を参照）。

【0274】

CGDDの、細胞株における安定した発現は、哺乳動物系の組換えタンパク質の長期にわたる産生のために望ましい。例えばCGDDをコードする配列を細胞株に形質転換するために、発現ベクター類と、同じベクター上の或いは別のベクター上の選択可能マーカー遺伝子とを用い得る。用いる発現ベクターは、ウイルス起源の複製要素、および／または内因性の発現エレメント群を持ち得る。ベクターの導入後、選択培地に移す前に強化培地で約1～2日間、細胞を増殖させ得る。選択可能マーカーの目的は選択剤への抵抗性を与えることであり、選択可能マーカーの存在により、導入した配列をうまく発現するような細胞の成長および回収が可能となる。安定的に形質転換された細胞の耐性クローンは、その細胞型に適した組織培養技術を用いて増殖可能である。

10

【0275】

任意の数の選択系を用いて、形質転換細胞株を回収できる。限定するものではないがこのような選択系には、 tk^- 細胞のために用いられる単純ヘルペスウイルスのチミジンキナーゼ遺伝子と、 apr^- 細胞のために用いられる単純ヘルペスウイルスのアデニンホスホリボシルトランスフェラーゼ遺伝子がある（例えばWigler, M.他(1977) Cell 11:223-232; Lowy, I. 他(1980) Cell 22:817-823を参照）。また、選択の基礎として代謝拮抗物質、抗生物質あるいは除草剤への耐性を用いることができる。例えばdhfrはメトトレキセートに対する耐性を与え、neoはアミノグリコシッドネオマイシンおよびG-418に対する耐性を与え、alsはクロルスルフロンに対する耐性を、patはホスフィノトリシンアセチルトランスフェラーゼに対する耐性を各々与える（例えばWigler, M.他(1980) Proc. Natl. Acad. Sci. USA 77:35673570; Colbere-Garapin, F.他(1981) J. Mol. Biol. 150:1-14を参照）。この他の選択可能な遺伝子（例えばtrpBおよびhisD、これらは代謝物のための、細胞の必要条件を改変）も、文献に記載がある（例えばHartman, S.C.およびR.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051を参照）。可視マーカー類、例えばアントシアニンや、緑色蛍光タンパク質（GFP; Clontech）、 β グルクロニダーゼおよびその基質である β グルクロニド、またはルシフェラーゼおよびその基質であるルシフェリンを用い得る。これらのマーカーを用いて、トランスフォーマントを同定するだけでなく、特定のベクター系に起因する一過性あるいは安定したタンパク質発現を定量し得る（例えばRhodes, C.A. (1995) Methods Mol. Biol. 55:121-131を参照）。

20

30

【0276】

マーカー遺伝子発現の有無によって目的の遺伝子の存在が示唆されても、その遺伝子の存在および発現の確認が必要な場合もある。例えばCGDDをコードする配列が或るマーカー遺伝子配列の中に挿入された場合、CGDDをコードする配列群を有する形質転換された細胞群は、マーカー遺伝子機能の欠落により同定できる。または、CGDDをコードする1配列とタンデムに、或るマーカー遺伝子を、単一プロモーターの制御下で配置することもできる。誘導または選択に応答したマーカー遺伝子の発現は通常、タンデム遺伝子の発現も示す。

【0277】

CGDDをコードする核酸配列を持ち、CGDDを発現する宿主細胞は一般に、当業者に周知の種々の方法を用いて特定できる。これらの方法には、DNA-DNA或いはDNA-RNAハイブリダイゼーションや、PCR法、核酸或いはタンパク質の検出及び／または数量化のための膜系、溶液ベース、或いはチップベースの技術を含むタンパク質生物学的試験法または免疫学的アッセイが含まれるが、これらに限定されるものではない。

40

【0278】

CGDDの発現の検出及び計測のための、特異的なポリクローナル抗体またはモノクローナル抗体のどちらかを用いる免疫学的方法は、当分野で既知である。このような技法の例には、酵素に結合した免疫吸着剤検定法（ELISA）、ラジオイムノアッセイ（RIA）、蛍光活性化細胞選別（FACS、フローサイトメトリー）などがある。CGDD上の2つの非干渉エピト

50

ープに反応するモノクローナル抗体群を用いた、2部位のモノクローナルベース免疫アッセイ (two-site, monoclonal-based immunoassay) が好ましいが、競合結合試験を用いることもできる。これらのアッセイおよび他のアッセイは、当分野で周知である (例えば Hampton, R.他(1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect.IV; Coligan, J.E.他(1997) Current Protocols in Immunology, Greene Publishing Associates and Wiley-Interscience, New York NY; および Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJを参照)。

【0279】

多岐にわたる標識方法および抱合方法が当業者に知られており、様々な核酸アッセイおよびアミノ酸アッセイにこれらの方法を用い得る。CGDDをコードするポリヌクレオチドに関連する配列を検出するための、標識されたハイブリダイゼーションプローブ或いはPCRプローブを生成する方法には、オリゴ標識化、ニックトランスレーション、エンドラベリング (末端標識化)、または標識されたヌクレオチドを用いるPCR増幅が含まれる。または、CGDDをコードする配列、またはその任意の断片を、mRNAプローブを生成するためのベクターにクローニングすることも可能である。このようなベクターは、当分野において知られており、市販もされており、T7、T3またはSP6などの好適なRNAポリメラーゼおよび標識されたヌクレオチドを加えて、*in vitro*でRNAプローブの合成に用いることができる。このような方法は、例えばAmersham Pharmacia Biotech、Promega (Madison WI)、U.S. Biochemicalなどから市販されている種々のキットを用いて実行することができる。検出を容易にするために用い得る好適なレポーター分子あるいは標識には、基質、補助因子、インヒビター、磁気粒子のほか、放射性核種、酵素、蛍光剤、化学発光剤、発色剤などがある。

【0280】

CGDDをコードするヌクレオチド配列で形質転換された宿主細胞は、細胞培地でのこのタンパク質の発現および回収に好適な条件下で培養される。形質転換細胞から製造されたタンパク質が分泌されるか細胞内に留まるかは、使用される配列、ベクター、あるいはその両者に依存する。CGDDをコードするポリヌクレオチドを持つ発現ベクターは、原核細胞膜または真核細胞膜を通してのCGDDの分泌を誘導するシグナル配列群を持つように設計できることは、当業者には理解されよう。

【0281】

更に、宿主細胞株の選択は、挿入した配列の発現をモジュレートする能力、または発現したタンパク質を所望の形に処理する能力によって行い得る。限定するものではないがこのようなポリペプチドの修飾には、アセチル化、カルボキシル化、グリコシル化、リン酸化、脂質化およびアシル化がある。タンパク質の「プレプロ」または「プロ」形を切断する翻訳後のプロセッシングを利用して、タンパク質の標的への誘導、折りたたみ、および/または活性を特定することが可能である。翻訳後の活性のための、特定の細胞装置および特徴的な機序を持つ、種々の宿主細胞 (例えばCHO、HeLa、MDCK、HEK293、WI38など) は、American Type Culture Collection (ATCC, Manassas, VA) から入手可能であり、外来タンパク質の正しい修飾およびプロセッシングを確実にするように選択し得る。

【0282】

CGDDをコードする、自然のあるいは修飾された、または組換えの核酸配列を、或る異種配列に結合させることにより、本発明の別の実施態様では、上記した任意の宿主系内で、1融合タンパク質の翻訳を生じ得る。例えば、市販の抗体によって認識できる異種部分を含むキメラCGDDタンパク質は、CGDDの活性のインヒビターに対するペプチドライブラリのスクリーニングを促進し得る。また、異種タンパク質部分および異種ペプチド部分も、市販されている親和性マトリックスを用いて融合タンパク質の精製を促進し得る。限定されるものではないがこのような部分には、グルタチオンSトランスフェラーゼ (GST)、マルトース結合タンパク質 (MBP)、チオレドキシン (Trx)、カルモジュリン結合ペプチド (CBP)、6-His、FLAG、c-myc、赤血球凝集素 (HA) がある。GSTは固定化グルタチオン上で、MBPはマルトース上で、Trxはフェニルアルシンオキシド上で、CBPはカルモジュリン上

10

20

30

40

50

で、そして6-Hisは金属キレート樹脂上で、同族の融合タンパク質の精製を可能にする。FLAG、c-mycおよび赤血球凝集素（HA）は、これらのエピトープ標識を特異的に認識する市販のモノクローナル抗体およびポリクローナル抗体を用いた、融合タンパク質の免疫親和性精製を可能にする。また、CGDDをコードする配列と異種タンパク質配列との間に、タンパク質分解切断部位を融合タンパク質が持つように遺伝子操作すると、CGDDが、精製後に異種部分から切断され得る。融合タンパク質の発現および精製方法は、前出のAusubel（1995）10章に記載されている。市販されている種々のキットを用いて融合タンパク質の発現および精製を促進することもできる。

【0283】

放射標識CGDDの合成は、本発明の別の実施態様で、TNTウサギ網状赤血球可溶化液またはコムギ胚芽抽出系(Promega)を用いてin vitroで可能である。これらの系は、T7、T3またはSP6プロモーターに機能的に連結したタンパク質コード配列群の、転写と翻訳とを共役させる。翻訳は、例えば³⁵Sメチオニンのような放射能標識したアミノ酸前駆体の存在下で起こる。

10

【0284】

本発明のCGDDまたはその断片を用いて、CGDDに特異結合する化合物をスクリーニングできる。CGDDへの特異的な結合のスクリーニングには、少なくとも1つ、または複数の試験化合物を用いる。試験化合物としては、抗体、オリゴヌクレオチド、タンパク質（例えば受容体）または小分子が挙げられる。

【0285】

或る実施態様では、このように同定された化合物は、CGDDの天然リガンドに密接に関連し、例えばリガンドやその断片であり、または天然基質や、構造的または機能的な擬態物質（mimetic）、あるいは自然結合パートナーである（例えばColigan, J.E.他(1991) Current Protocols in Immunology 1(2):5章を参照）。同様に、この化合物は、CGDDが結合する天然受容体に、或いは例えばリガンド結合部位など少なくともこの受容体の或る断片に、密接に関連し得る。何れの場合も、既知の技術を用いてこの化合物を合理的に設計することができる。或る実施例では、これら化合物のためのスクリーニングには、分泌タンパク質として、あるいは細胞膜上でCGDDを発現する、好適な細胞群の作製が含まれる。好適な細胞には、哺乳類、酵母、ショウジョウバエ、または大腸菌からの細胞が含まれる。CGDDを発現する細胞、またはCGDDを含有する細胞膜断片を試験化合物と接触させて、結合や、CGDDまたは該化合物の何れかの、刺激または活性の阻害を分析する。

20

30

【0286】

或るアッセイは、単に試験化合物をポリペプチドに実験的に結合させ、フルオロフォア、放射性同位体、酵素抱合体、または他の検出可能な標識により、その結合を検出できる。例えば、このアッセイは、少なくとも1つの試験化合物を、溶液中の、あるいは固体支持物に固定されたCGDDと混合するステップと、CGDDとこの化合物との結合を検出するステップを含み得る。別法では、標識された競合物の存在下での試験化合物の結合の検出および測定を行うことができる。更にこのアッセイは、無細胞再構成標本、化学ライブラリ、または、天然産物の混合物を用いて実施でき、試験化合物（群）は、溶液中で遊離させるか固体支持体に固定する。

40

【0287】

本発明のCGDDまたはその断片を用いて、CGDDの活性をモジュレートする化合物をスクリーニングし得る。このような化合物としては、アゴニスト、アンタゴニスト、部分的アゴニストまたは逆アゴニストなどが含まれ得る。一実施態様では、CGDDが少なくとも1つの試験化合物と混合される、CGDD活性が許容される諸条件下で或るアッセイを実施し、試験化合物の存在下でのCGDDの活性を試験化合物不在下でのCGDDの活性と比較する。試験化合物の存在下でのCGDDの活性の変化は、CGDDの活性をモジュレートする化合物の存在を示唆する。別法では、試験化合物を、CGDDの活性に適した条件下で、CGDDを有するin vitro系すなわち無細胞系と混合してアッセイを実施する。CGDDの活性を調節する試験化合物は、これらアッセイの何れにおいても間接的に調節する場合があり、その際は試験化合物と直接

50

接触する必要がない。少なくとも1つ、または複数の試験化合物をスクリーニングすることができる。

【0288】

別の実施例では、CGDDまたはその哺乳動物相同体をコードするポリヌクレオチドを、胚性幹細胞（ES細胞）における相同組換えを用いて動物モデル系内で「ノックアウト」する。このような技術は当技術分野において周知であり、ヒト疾患動物モデルの作製に有用である（米国特許第5,175,383号および第5,767,337号などを参照）。例えば129/SvJ細胞系などのマウスES細胞は初期のマウス胚に由来し、培地で増殖させることができる。このES細胞は、ネオマイシンホスホトランスフェラーゼ遺伝子（neo: Capecchi, M.R. (1989) Science 244:1288-1292）などのマーカー遺伝子で破壊した、目的の遺伝子を持つベクターで形質転換する。このベクターは、相同組換えにより、宿主ゲノムの対応する領域に組込まれる。或いは、Cre-loxP系を用いて相同組換えを行い、組織特異的または発生段階特異的に目的遺伝子をノックアウトする（Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U.他(1997) Nucleic Acids Res.25:4323-4330）。形質転換したES細胞を同定し、例えばC57BL/6マウス株などから採取したマウス細胞胚盤胞に微量注入する。これらの胚盤胞を偽妊娠メスに外科的に導入し、得られるキメラ子孫の遺伝形質を特定し、これらを交配させてヘテロ接合性系またはホモ接合性系を作製する。このようにして作製した遺伝子組換え動物は、潜在的な治療薬や毒性薬物で検査されうる。

10

【0289】

CGDDをコードするポリヌクレオチド類はまた、*in vitro*で、ヒト胚盤胞由来のES細胞において操作し得る。ヒトES細胞は、内胚葉、中胚葉および外胚葉の細胞タイプを含む、少なくとも8つの別々の細胞系統に分化する可能性を有する。これらの細胞系統は、例えば神経細胞、造血系統および心筋細胞に分化する（Thomson, J.A.他(1998) Science 282:1145-1147）。

20

【0290】

CGDDをコードするポリヌクレオチドを用いて、ヒト疾患をモデルとした「ノックイン」ヒト化動物（ブタ）または遺伝子組換え動物（マウスまたはラット）を作製することも可能である。CGDDをコードするポリヌクレオチドの或る領域を、ノックイン技術を用いて動物ES細胞に注入し、注入した配列を動物細胞ゲノムに組込ませる。形質転換細胞を胞胚に注入し、胞胚を上記のように移植する。遺伝子組換え子孫または近交系について試験し、潜在的医薬品を用いて処理し、ヒトの疾患の治療に関する情報を得る。または、CGDDを例えば乳汁内に分泌するなど過剰に発現する哺乳動物近交系は、便利なタンパク質源となり得る（Janne, J. 他 (1998) Biotechnol. Annu. Rev. 4:55-74）。

30

【0291】

（治療）

CGDDの領域群と細胞成長、分化および細胞死関連タンパク質との間には、例えば配列及びモチーフの文脈における、化学的および構造的類似性が存在する。またCGDDを発現する組織の例は乳癌、PBMC細胞、脳带状組織であり、また表6を見られたい。このように、CGDDは、癌など細胞増殖異常、発達障害、神経疾患、自己免疫／炎症疾患、生殖障害、および胎盤障害で役割を果たすと考えられる。CGDDの発現若しくは活性の増大に関連する疾患の治療においては、CGDDの発現または活性を低下させることが望ましい。CGDDの発現または活性の低下に関連する疾患の治療においては、CGDDの発現または活性を亢進させることが望ましい。

40

【0292】

従って一実施態様において、CGDDの発現または活性の低下に関連した疾患の治療または予防のために、被験者にCGDDまたはその断片や誘導体を投与し得る。限定するものではないが、このような疾患には細胞増殖異常、発達障害、神経疾患、自己免疫／炎症疾患、生殖障害、胎盤障害が含まれ、細胞増殖異常には日光角化症、動脈硬化、アテローム硬化、滑液包炎、硬変、肝炎、混合性結合組織病（MCTD）、骨髓線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症、並びに、腺癌、白血病、リンパ腫、黒色腫

50

、骨髓腫、肉腫、および奇形癌など癌、具体的には、副腎、膀胱、骨、骨髓、脳、乳房、頸部、胆嚢、神経節、消化管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、および子宮の癌が含まれる。発達障害には尿細管性アシドーシス、貧血、クッシング症候群、軟骨形成不全性小人症、デュシェンヌ／ベッカー型筋ジストロフィー、癲癇、性腺形成異常、WAGR症候群（ウィルムス腫瘍、無虹彩症、尿生殖器異常、精神遅滞）、スミス－マジエニス症候群（Smith-Magenis syndrome）、骨髓異形成症候群、遺伝性粘膜上皮異形成、遺伝性角皮症や、シャルコーマリーツース病および神経線維腫症などの遺伝性神経病、甲状腺機能低下症、水頭症や、Syndenham舞蹈病（Syndenham's chorea）および脳性小児麻痺などの発作障害、二分脊椎、無脳症、頭蓋脊椎披裂、先天性緑内障、白内障、感音難聴が含まれる。神経疾患には、癲癇、虚血性脳血管障害、脳卒中、脳腫瘍、アルツハイマー病、ピック病、ハンチントン病、痴呆、パーキンソン病およびその他の錐体外路障害、筋萎縮性側索硬化およびその他の運動ニューロン障害、進行性神経性筋萎縮症、網膜色素変性症（色素性網膜炎）、遺伝性運動失調、多発性硬化症および他の脱髄疾患、細菌性およびウイルス性髄膜炎、脳膿瘍、硬膜下膿瘍、硬膜外膿瘍、化膿性頭蓋内血栓性静脈炎、脊髄炎および神経根炎、ウイルス性中枢神経系疾患と、プリオン病（クールー、クロイツフェルト－ヤコブ病、およびGerstmann-Straussler-Scheinker症候群を含む）、致死性家族性不眠症、神経系性栄養病および代謝病、神経線維腫症、結節硬化症、小脳網膜血管腫症（cerebelloretinal hemangioblastomatosis）、脳3叉神経血管症候群、ダウン症を含む中枢神経系性精神遅滞および他の発達障害、脳性麻痺、神経骨格異常症、自律神経系障害、脳神経疾患、脊髄疾患、筋ジストロフィー他の神経筋障害、末梢神経疾患、皮膚筋炎および多発性筋炎と、遺伝性、代謝性、内分泌性、および中毒性ミオパシーと、重症筋無力症、周期性四肢麻痺、精神障害（気分性、不安性の障害、統合失調症／分裂病）、季節性感情障害（SAD）、静座不能、健忘症、緊張病、糖尿病性ニューロパシー、遅発性ジスキネジア、ジストニー、パラノイド精神病、帯状疱疹後神経痛、トゥーレット病、進行性核上麻痺、大脳皮質基底核変性（corticobasal degeneration）、および家族性前頭側頭型痴呆とが含まれる。自己免疫／炎症疾患には、後天性免疫不全症候群（AIDS）、アジソン病（慢性原発性副腎機能不全）、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫性多腺性内分泌カンジダ症外胚葉ジストロフィー（APECED）、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、糖尿病、肺気腫、リンパ球傷害因子性偶発性リンパ球減少症、新生児溶血性疾患（胎児赤芽球症）、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性腸症候群、多発性硬化症、重症筋無力症、心筋または心膜炎、骨関節炎、骨粗しょう症、膵炎、多発性筋炎、乾癬、ライター症候群、リウマチ様関節炎、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性強皮症、血小板減少性紫斑病、潰瘍性大腸炎、ぶどう膜炎、ウェルナー症候群、癌合併症、血液透析、体外循環、ウイルス感染症、細菌感染症、真菌感染症、寄生虫感染症、原虫感染症、蠕虫感染症、および外傷が含まれる。生殖障害には、プロラクチン産生の障害、不妊（例えば卵管疾患、排卵不良、子宮内膜症、性周期の途絶、月経周期の途絶、多嚢胞性卵巣症候群、卵巣過剰刺激症候群、子宮内膜腫瘍や卵巣腫瘍、子宮筋腫、自己免疫疾患、異所性妊娠、奇形発生、乳癌、乳房線維嚢胞病、乳漏症、精子形成の途絶、異常精子生理、精巣癌、前立腺癌、良性前立腺肥大、前立腺炎、ペーロニー病、インポテンス、男性乳癌、女性化乳房、高ゴナドトロピン性腺機能低下症、低ゴナドトロピン性腺機能低下症、仮性半陰陽、無精子症、早発卵巣不全、アクロシン欠損症、晩発思春期、逆行性射精、無射精、血管芽腫、嚢胞・クロム親和細胞腫（cystsphaeochromocytomas）、傍神経節腫、精巣上体の嚢胞腺腫、および内リンパ嚢腫瘍）を含む。胎盤障害には、子癰前症、絨毛癌、胎盤早期剥離、前置胎盤、胎盤梗塞やmaternal floor infarction、癒着胎盤、侵入胎盤（increate）および穿通胎盤、絨毛膜外性胎盤（extrachorial placentas）、絨毛血管腫（chorangioma）、絨毛血管分布過多（chorangiosis）、慢性絨毛炎（villitis）、胎盤絨毛浮腫（pla

cental villous endema)、終末絨毛広範性線維症(widespread fibrosis of the terminal villi)、絨毛間腔血栓(intervillous thrombi)、出血性血管炎、新生児溶血性疾患(胎児赤芽球症)、非免疫性胎児水腫を含む。

【0293】

別の実施態様では、限定するものではないが上に列記した疾患を含む、CGDDの発現または活性の低下に関連した疾患の治療または予防のために、CGDDまたはその断片や誘導体を発現し得るベクターを被験者に投与し得る。

【0294】

更に別の実施態様では、限定するものではないが上に列記した疾患を含む、CGDDの発現または活性の低下に関連した疾患の治療または予防のために、実質的に精製されたCGDDを有する組成物を好適な医薬用キャリアと共に被験者に投与し得る。 10

【0295】

更に別の実施態様では、限定するものではないが上に列記した疾患を含む、CGDDの発現または活性の低下に関連した疾患の治療または予防のために、CGDDの活性を調節するアゴニストを被験者に投与し得る。

【0296】

更なる実施態様では、CGDDの発現または活性の増大に関連した疾患の治療または予防のために、被験者にCGDDのアンタゴニストを投与し得る。限定するものではないが、このような疾患の例には、上記した癌など細胞増殖異常、発達障害、神経疾患、自己免疫/炎症の疾患、生殖障害、および胎盤障害が含まれる。一実施態様では、CGDDと特異的に結合する抗体が直接アンタゴニストとして、或るいはCGDDを発現する細胞または組織に薬剤を運ぶターゲティング機構或いは送達機構として間接的に用いられ得る。 20

【0297】

別の実施態様では、CGDDをコードするポリヌクレオチドの相補配列を発現するベクターを被験者に投与して、限定するものではないが上記した疾患を含む、CGDDの発現または活性の増大に関連した疾患の治療または予防も可能である。

【0298】

別の実施態様では、本発明の任意のタンパク質、アンタゴニスト、抗体、アゴニスト、相補配列、またはベクターを、別の好適な治療薬と組合せて投与することもできる。併用療法で用いる好適な治療薬は、当業者が従来の医薬原理に従って選択し得る。治療薬と組合せることにより、上記した種々の疾患の治療または予防に相乗効果をもたらし得る。この方法により、少量の各薬物で医薬効果をあげることが可能となり、それによって副作用の可能性を低減し得る。 30

【0299】

CGDDのアンタゴニストは、当分野で一般に既知の方法を用いて製造し得る。具体的には、精製CGDDを用いて抗体を作るか、治療薬のライブラリをスクリーニングして、CGDDと特異結合するものを同定し得る。CGDDへの抗体も、当分野で周知の方法を用いて産生され得る。限定するものではないがこのような抗体には、ポリクローナル抗体、モノクローナル抗体、キメラ抗体、一本鎖抗体、Fab断片、およびFab発現ライブラリによって作られた断片が含まれ得る。中和抗体(すなわち二量体の形成を阻害する抗体)は通常、治療用に好適である。一本鎖抗体(例えば由来はラクダまたはラマ)は強力な酵素阻害剤である可能性があり、ペプチド擬態物質の設計において利点を持つようであり、免疫吸着剤とバイオセンサーとの開発においても利点を持つようである(Muyldermans, S. (2001) J. Biotechnology 1, 74:277-302)。 40

【0300】

CGDD、若しくは免疫原性の特性を備えるその任意の断片またはオリゴペプチドの注入によって、抗体の産生のために、ヤギ、ウサギ、ラット、マウス、ラクダ、ヒトコブラクダ、ラマ、ヒトなどを含む種々の宿主が免疫化され得る。宿主の種に応じて、種々のアジュバントを用いて免疫応答を高めることもできる。限定するものではないがこのようなアジュバントには、フロイントアジュバントと、水酸化アルミニウムなどのミネラルゲルアジュ 50

バントと、リゾレシチン、プルロニックポリオール、ポリアニオン、ペプチド、油性乳剤、スカシガイのヘモシアニン (KLH)、ジニトロフェノールなどの界面活性剤とがある。ヒトに用いられるアジュバントの中では、BCG (カルメット - ゲラン杆菌) およびコリネバクテリウム - パルバム (*Corynebacterium parvum*) が特に好ましい。

【0301】

CGDDに対する抗体を誘導するために用いるオリゴペプチド、ペプチドまたは断片は、少なくとも約5個のアミノ酸からなるアミノ酸配列を持つものが好ましく、一般的には約10個以上のアミノ酸からなるものとなる。これらのオリゴペプチド、ペプチドまたは断片は、天然タンパク質のアミノ酸配列の一部と同一であることが望ましい。CGDDアミノ酸類の短いストレッチ群は、KLHなど他のタンパク質の配列と融合され、このキメラ分子に対する抗体群が産生され得る。

10

【0302】

CGDDに対するモノクローナル抗体は、培地内の連続した細胞株によって抗体分子群を産生する、任意の技術を用いて作製できる。限定するものではないがこのような技術には、ハイブリドーマ技術、ヒトB細胞ハイブリドーマ技術、およびEBV-ハイブリドーマ技術がある (例えばKohler, G.他(1975) *Nature* 256:495-497; Kozbor, D.他(1985) *J. Immunol. Methods* 81:31-42; Cote, R.J.他 (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; および Cole, S.P.他 (1984) *Mol. Cell Biol.* 62:109-120を参照)。

【0303】

更に、「キメラ抗体」作製のために発達した、ヒト抗体遺伝子にマウス抗体遺伝子をスプライシングするなどの技術が、好適な抗原特異性および生物学的活性を備える分子を得るために用いられる (例えばMorrison, S.L.他(1984) *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. 他 (1984) *Nature* 312:604-608; および Takeda, S.他 (1985) *Nature* 314:452-454を参照)。または、CGDD特異の一本鎖抗体を、当分野で既知の方法を用いて、一本鎖抗体の産生のための記載された技術を適用して生成し得る。関連した特異性を有するがイディオタイプ組成が異なるような抗体を、ランダムな組合せの免疫グロブリンライブラリ類からチェーンシャッフリングによって産生することもできる (例えばBurton D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137を参照)。

20

【0304】

抗体類の産生は、リンパ球集団における *in vivo* 産生の誘導によって、或いは文献に開示されているように非常に特異的な結合試薬の免疫グロブリンのライブラリ類またはパネル類のスクリーニングによっても行い得る (例えばOrlandi, R.他(1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter, G.他 (1991) *Nature* 349:293-299を参照)。

30

【0305】

CGDDに対する特異結合部位を持つ抗体断片をも産生し得る。例えば、限定するものではないが、このような断片には、抗体分子のペプシン消化によって作製される $F(ab')_2$ 断片と、 $F(ab')_2$ 断片のジスルフィド架橋を還元することによって作製されるFab断片とがある。あるいは、Fab発現ライブラリを作製することによって、所望の特異性を持つモノクローナルFab断片を迅速且つ容易に同定することが可能となる (例えばHuse, W.D.他(1989) *Science* 246:1275-1281を参照)。

40

【0306】

種々の免疫学的検定 (イムノアッセイ) を用いてスクリーニングすることにより、所望の特異性を有する抗体を同定し得る。確立された特異性を有するポリクローナル抗体またはモノクローナル抗体の何れかを用いる免疫放射定量測定法または競合結合試験に対する数々のプロトコルは、本技術分野において公知である。このようなイムノアッセイは通常、CGDDとその特異抗体との間の複合体形成の計測を含む。2つの非干渉性CGDDエピトープに対して反応性を持つモノクローナル抗体群を用いる、2部位モノクローナルベースのイムノアッセイが一般に利用されるが、競合結合試験も利用できる (Pound、前出)。

【0307】

CGDDに対する抗体の親和性を、ラジオイムノアッセイ技術と共にScatchard分析などの様

50

々な方法を用いて算定する。CGDD抗体複合体のモル濃度を平衡状態の下で遊離抗体と遊離抗原のモル濃度で除して得られる結合定数 K_a が親和性を表す。CGDDエピトープ多数に対してポリクローナル抗体類は親和性が不均一であり、或るポリクローナル抗体製剤に関して判定した K_a は、CGDDに対する抗体群の平均の親和性または結合活性を表す。特定CGDDエピトープに対してモノクローナル抗体は単一特異的であり、モノクローナル抗体の或る製剤について判定した K_a は、親和性の真の測定値を表す。CGDD-抗体複合体が激しい操作に耐えなければならないイムノアッセイには、 K_a 値が $10^9 \sim 10^{12} \text{ L/mol}$ の高親和性抗体製剤を用いるのが好ましい。CGDDが抗体から最終的に（好ましくは活性化状態で）解離する必要がある免疫精製（immunopurification）および類似の処理には、 K_a 値が $10^6 \sim 10^7 \text{ L/mol}$ の低親和性抗体製剤を用いるのが好ましい（Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington, DC; Liddell, J. E. および Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY）。

10

【0308】

ポリクローナル抗体製剤の抗体価および結合活性を更に評価して、後に使う或る適用例に対するこのような製剤の品質および適性を決定することができる。例えば、少なくとも $1 \sim 2 \text{ mg/ml}$ の特異的な抗体、好ましくは $5 \sim 10 \text{ mg/ml}$ の特異的な抗体を含むポリクローナル抗体製剤は一般に、CGDD抗体複合体を沈殿させなければならない処理に用いられる。抗体の特異性、抗体価、結合活性、様々な適用例における抗体の品質や使用に対する指針については、一般に入手可能である（例えば前出のCatty、同Coligan 他を参照）。

20

【0309】

CGDDをコードするポリヌクレオチド、またはその任意の断片や相補配列が、本発明の別の実施態様では治療目的で使用され得る。ある態様では、CGDDをコードする遺伝子のコーディング領域や調節領域に相補的な配列やアンチセンス分子（DNA、RNA、PNA、または修飾したオリゴヌクレオチド）を設計して遺伝子発現を変更することができる。CGDDをコードする配列の制御領域、またはコード領域に沿った、さまざまな位置から、アンチセンスオリゴヌクレオチドまたはより大きな断片を、このような当分野で周知の技術で設計可能である（例えばAgrawal, S., 編集 (1996) Antisense Therapeutics, Humana Press Inc., Tottowa NJを参照）。

30

【0310】

治療に用いる場合、アンチセンス配列を適切な標的細胞に導入するのに好適な、任意の遺伝子送達系を用いることができる。アンチセンス配列は、転写時に標的タンパク質をコードする細胞配列の少なくとも一部に相補的な配列を作製する発現プラスミドの形で、細胞内に送達し得る（例えばSlater, J.E. 他(1998) *J. Allergy Clin. Immunol.* 102(3):469-475; およびScanlon, K.J. 他(1995) 9(13):1288-1296を参照）。アンチセンス配列はまた、例えばレトロウイルスやアデノ随伴ウイルスベクターなどのウイルスベクターを用いて細胞内に導入することもできる（例えばMiller, A.D. (1990) *Blood* 76:271、前出のAusubel, Uckert, W. およびW. Walther (1994) *Pharmacol. Ther.* 63 (3) :323-347を参照）。その他の遺伝子送達機構には、リポソーム系、人工的なウイルスエンベロープおよび当分野で公知のその他のシステムが含まれる（例えばRossi, J.J. (1995) *Br. Med. Bull.* 51 (1) :217-225; Boado, R.J. 他(1998) *J. Pharm. Sci.* 87(11):1308-1315、Morris, M. C. 他 (1997) *Nucleic Acids Res.* 25(14):2730-2736を参照）。

40

【0311】

CGDDをコードするポリヌクレオチドを、本発明の別の実施態様では、体細胞若しくは生殖細胞の遺伝子治療に用いることが可能である。遺伝子治療により、(i) 遺伝子欠損症を治療し（例えばX染色体連鎖遺伝（Cavazzana-Calvo, M. 他(2000) *Science* 288:669-672）により特徴付けられる重症複合型免疫不全(SCID)-X1病の場合）、先天性アデノシンデアミナーゼ (ADA) 欠損症に関連する重症複合型免疫不全症候群（Blaese, R.M. 他(1995) *Science* 270:475-480; Bordignon, C. 他(1995) *Science* 270:470-475）、嚢胞性繊維症

50

(Zabner, J.他(1993) Cell 75:207-216; Crystal, R.G.他(1995) Hum.Gene Therapy 6:643-666; Crystal, R.G.他(1995) Hum.Gene Therapy 6:667-703)、サラセミア(thalassaemia)、家族性高コレステロール血症や、第8因子若しくは第9因子欠損に起因する血友病(Crystal, R.G. (1995) Science 270:404-410、Verma, I.M. および N. Somia (1997) Nature 389:239-242)、(ii) 条件的致死性遺伝子産物を発現させ(例えば制御不能な細胞増殖に起因する癌の場合)、(iii) 細胞内の寄生生物、例えばヒト免疫不全ウイルス(HIV) (Baltimore, D. (1988) Nature 335:395-396、Poeschla, E.他(1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399) などヒトレトロウイルス、B型若しくはC型肝炎ウイルス(HBV、HCV)、Candida albicansおよびParacoccidioides brasiliensis等の寄生真菌、並びに熱帯熱マラリア原虫およびクルーズトリパノソーマ等の寄生原虫に対する防御機能を有するタンパク質を発現させることができる。CGDDの発現若しくは調節に必要な遺伝子の欠損が疾患を引き起こす場合、導入した細胞の好適な集団からCGDDを発現させて、遺伝子欠損によって起こる症状の発現を緩和することが可能である。

10

【0312】

本発明の更なる実施例では、CGDDの欠損による疾患や障害を、CGDDをコードする哺乳動物発現ベクターを作製して、これらのベクターを機械的手段によってCGDD欠損細胞に導入することによって治療する。in vivoあるいはex vitroの細胞に用いる機械的導入技術には、(i) 個々の細胞内への直接的なDNA微量注射法、(ii) 遺伝子銃、(iii) リポソームを介した形質移入、(iv) 受容体を介した遺伝子導入、および(v) DNAトランスポゾンの使用がある(Morgan, R.A. および W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217、Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. および H. Recipon (1998) Curr. Opin. Biotechnol. 9:445-450)。

20

【0313】

CGDDの発現に有効であり得る発現ベクターの例として、限定するものではないがpcDNA 3.1、EpiTag、pRcCMV2、pREP、pVAX、pCRII-TOPOTAベクター(Invitrogen, Carlsbad CA)、pCMV-Script、pCMV-Tag、PEGSH/PERV(Stratagene, La Jolla CA) およびPTET-OFF、PTET-ON、PTRE2、PTRE2-LUC、PTK-HYG(Clontech, Palo Alto CA) が挙げられる。CGDDを発現させるために、(i) 構成的に活性なプロモーター(例えば、サイトメガロウイルス(CMV)、ラウス肉腫ウイルス(RSV)、SV40ウイルス、チミジンキナーゼ(TK)、若しくはβ-アクチンの遺伝子など)、(ii) 誘導性プロモーター(例えば、市販されているT-REXプラスミド(Invitrogen)に含まれている、テトラサイクリン調節性プロモーター(Gossen, M. および H. Bujard (1992) Proc. Natl. Acad. Sci. U.S.A. 89:5547-5551; Gossen, M. 他 (1995) Science 268:1766-1769; Rossi, F.M.V. および H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456))、エクジソン誘導性プロモーター(市販されているプラスミドPVGRXRおよびPINDに含まれている:Invitrogen)、FK506/ラパマイシン誘導性プロモーター、またはRU486/ミフェプリストン誘導性プロモーター(Rossi, F.M.V. および H.M. Blau, 前出)、または(iii) 正常な個体に由来する、CGDDをコードする内因性遺伝子の天然のプロモーター若しくは組織特異的プロモーターを用いることが可能である。

30

【0314】

市販のリポソーム形質転換キット(例えばInvitrogen社のPerFect Lipid Transfection Kit)を用いれば、当業者は実験の各パラメータを最適化する努力をさほど要さず、ポリヌクレオチド群を、培養中の標的細胞群に送達し得る。別法では、リン酸カルシウム法(Graham, F.L. および A.J. Eb (1973) Virology 52:456-467)若しくは電気穿孔法(Neuman, B. 他(1982) EMBO J. 1:841-845)によって、形質転換が実施される。初代培養細胞にDNAを導入するためには、標準化された哺乳類の形質移入プロトコルの修正が必要である。

40

【0315】

本発明の別の実施態様では、CGDDの発現に関連する遺伝子欠損によって起こる疾患や障害は、(i) レトロウイルス末端反復配列(LTR)プロモーター若しくは独立したプロモーターの調節の下でCGDDをコードするポリヌクレオチドと、(ii) 好適なRNAパッケージング

50

シグナルと、(iii) 追加のレトロウイルス・シス作用性RNA配列及び効率的なベクターの増殖に必要なコード配列を伴うRev応答性エレメント (RRE) と、からなるレトロウイルスベクターを作製して治療することができる。レトロウイルスベクター (例えばPFBおよびPFBNEO) はStratagene社から市販されており、刊行データ (Riviere, I. 他 (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737) に基づく。上記データを引用することを以て本明細書の一部とする。このベクターは、好適なベクター産生細胞株 (VPCL) において増殖される。VPCLは、各標的細胞上の受容体への親和性を持つエンベロープ遺伝子を、またはVS Vgなど汎親和性エンベロープタンパク質を発現する (Armentano, D. 他 (1987) J. Virol. 61:1647-1650; Bender, M.A. 他 (1987) J. Virol. 61:1639-1646; Adam, M.A. および A .D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. 他 (1998) J. Virol. 72:8463-84 10 71; Zufferey, R. 他 (1998) J. Virol. 72:9873-9880)。Riggに付与された米国特許第5,910,434号 (「Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant」) において、レトロウイルスパッケージング細胞株群を得る方法が開示されており、引用を以て本明細書の一部とする。レトロウイルスベクター類の繁殖や、細胞集団 (例えばCD4⁺ T細胞群) の形質導入、および形質導入した細胞群の患者への戻しは、遺伝子治療の分野では当業者に周知の手法であり、多数の文献に記載がある (Ranga, U. 他 (1997) J. Virol. 71:7020-7029; Bauer, G. 他 (1 997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. 他 (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2 20 290)。

【0316】

または、CGDDの発現に関連する1つ以上の遺伝子異常を有する細胞に、アデノウイルス系遺伝子治療の送達系を用いて、CGDDをコードするポリヌクレオチドを送達する。アデノウイルス系ベクター類の作製およびパッケージングについては、当業者に周知である。複製欠損型アデノウイルスベクター類は、種々の免疫調節タンパク質をコードする遺伝子群を、無損傷の臍島内に導入する目的で多様に利用し得ることが証明された (Csete, M.E. 他 (1995) Transplantation 27:263-268)。潜在的に有用なアデノウイルスベクター類はArmentanoに付与された米国特許第5,707,618号 (「Adenovirus vectors for gene therapy」) に記載されており、引用することを以て本明細書の一部とする。アデノウイルスベクター類については、Antinozzi, P.A. 他 (1999) Annu. Rev. Nutr. 19:511-544 並びに、Verma 30 , I.M. および N. Somia (1997) Nature 18:389:239-242も参照されたい。両文献は、引用を以て本明細書の一部とする。

【0317】

または、CGDDの発現に関連する1つ以上の遺伝子異常を有する標的細胞に、ヘルペス系遺伝子治療の送達系を用いて、CGDDをコードするポリヌクレオチドを送達する。CGDDを単純疱疹ウイルス (HSV) が向性を持つ中枢神経細胞に導入するには、HSV系ベクターの利用が特に有用たり得る。ヘルペス系ベクター類の作製およびパッケージングは、当業者に公知である。或る複製適格性単純ヘルペスウイルス (HSV) 1型系のベクターが、或るレポーター遺伝子の、霊長類の眼への送達に用いられている (Liu, X. 他 (1999) Exp. Eye Res. 169:385-395)。HSV-1ウイルスベクターの作製についても、DeLucaに付与された米国特許 40 第5,804,413号 (「Herpes simplex virus strains for gene transfer」) に詳細に開示されており、該特許の引用を以て本明細書の一部とする。米国特許第5,804,413号は、ヒト遺伝子治療などの目的で、好適なプロモーターの制御下において細胞に導入される少なくとも1つの外来遺伝子を有するゲノムを持つ、組換えHSV d92の利用について記載する。上記特許はまた、ICP4、ICP27、およびICP22を欠失した組換えHSV系統の、作製および使用について開示している。HSVベクター類については、Goins, W.F. 他 (1999) J. Virol . 73:519-532 および Xu, H. 他 (1994) Dev. Biol. 163:152-161も参照されたい。両文献は、引用を以て本明細書の一部とする。クローン化したヘルペスウイルス配列群の操作、大ヘルペスウイルスゲノム (large herpesvirus genomes) の様々なセグメントを持つ複数のプラスミドを形質移入した後の組換えウイルスの産生、ヘルペスウイルスの成長およ 50

び増殖、並びに細胞群へのヘルペスウイルスの感染は、当業者に公知の技術である。

【0318】

または、CGDDをコードするポリヌクレオチド群を、或るαウイルス（正の一本鎖RNAウイルス）ベクターを用いて標的細胞群に送達する。プロトタイプのαウイルスであるセムリキ森林熱ウイルス（SFV）の生物学的研究が広範に行われており、遺伝子導入ベクター類は、SFVゲノムに基づく（Garoff, H. および K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469）。αウイルスRNAの複製中に、ウイルスのカプシドタンパク質群を通常はコードする、サブゲノムRNAが産生される。このサブゲノムRNAは、完全長ゲノムRNAよりも高レベルに複製するので、酵素活性（例えばプロテアーゼおよびポリメラーゼ）を持つウイルスタンパク質に比べてカプシドタンパク質群が過剰産生される。同様に、CGDDをコードする配列をαウイルスゲノムのカプシドをコードする領域に導入することによって、ベクター導入細胞において多数のCGDDをコードするRNAが産生され、高レベルでCGDDが合成される。通常、αウイルスの感染は、数日以内の細胞溶解を伴う。一方、シンドビスウイルス（SIN）の或る変異体を有するハムスター正常腎臓細胞（BHK-21）群が持続的な感染を確立する能力は、αウイルス類の溶解複製を、遺伝子治療に応用し得るように好適に改変可能であることを示唆する（Dryga, S.A.他(1997) Virology 228:74-83）。CGDDの、様々なタイプの細胞への導入を、αウイルスの広い宿主域が可能にする。或る集団における或るサブセットの細胞群の特異的形質導入は、形質導入前に細胞の選別を必要とし得る。αウイルスの感染性cDNAクローンの処置方法、αウイルスのcDNAおよびRNAの形質移入方法およびαウイルスの感染方法は、当業者に公知である。

10

20

【0319】

転写開始部位（transcription initiation site）由来のオリゴヌクレオチドを用いて遺伝子発現を阻害することも可能である。この位置は、例えばスタート部位(start site)から数えて約-10と約+10の間である。同様に、三重らせん塩基対の形成方法を用いて阻害が可能となる。三重らせん塩基対形成は、ポリメラーゼ、転写因子または調節分子の結合のために十分に開く、二重らせんの能力を阻害するので有用である。三重らせんDNAを用いる最近の治療の進歩については文献に記載がある（例えばGee, J.E.他(1994) in Huber, B.E. および B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, 163-177ページを参照）。また、相補配列またはアンチセンス分子を設計し、転写物がリボソームに結合するのを阻止することによってmRNAの翻訳をブロックし得る。

30

【0320】

リボザイム類は、酵素的RNA分子である。RNAの特異的切断を触媒するためにリボザイムを用い得る。リボザイム作用のメカニズムは、相補的標的RNAへのリボザイム分子の配列特異的ハイブリダイゼーションと、その後起こる内ヌクレオチド鎖切断に関与している。例えば、CGDDをコードする配列の、内ヌクレオチド分解切断を、人工的に作製されたハンマーヘッド型リボザイム分子が、特異的且つ効果的に触媒できる可能性がある。

【0321】

任意の潜在的RNA標的内の特異的リボザイム切断部位を先ず同定するため、GUA、GUU、GUC配列などリボザイム切断部位に対して標的分子をスキャンする。一度同定されると、切断部位を持つ標的遺伝子の領域に対応する15～20リボヌクレオチドの短いRNA配列が、そのオリゴヌクレオチドを機能不全にするような2次構造の特徴をもっていないかを評価することが可能になる。候補標的の適合性の評価も、リボヌクレアーゼ保護アッセイを用いて相補的オリゴヌクレオチド群とのハイブリダイゼーションへのアクセス可能性をテストすることによって行い得る。

40

【0322】

本発明の相補リボ核酸分子およびリボザイムは、核酸分子合成のために当分野でよく知られている任意の方法を用いて作製し得る。作製方法には、固相ホスホラミダイト化学合成など、オリゴヌクレオチドを化学的に合成する方法がある。または、CGDDをコードするDNA配列のin vitroおよびin vivo転写によって、RNA分子を産生し得る。このようなDNA配列

50

は、T7やSP6などの好適なRNAポリメラーゼプロモーターを用いて多様なベクター内に取り込むことが可能である。あるいは、相補的RNAを構成的あるいは誘導的に合成するこれらのcDNA構成物を、細胞株、細胞または組織内に導入することができる。

【0323】

細胞内の安定性を高め、半減期を長くするために、RNA分子を修飾し得る。限定するものではないが可能な修飾としては、分子の5'末端、3'末端、あるいはその両方において隣接配列群を追加することや、分子の主鎖内においてホスホジエステラーゼ結合ではなくホスホロチオエートまたは2'-O-メチルを使用することが含まれる。この概念は、本来はPNA群の産出におけるものであるが、これら全ての分子に拡大することができる。そのためには、内因性エンドヌクレアーゼによって容易には認識されない、アデニン、シチジン、グアニン、チミン、およびウリジンにアセチル、メチル、チオールおよび同様の修飾をしたものや、非従来型塩基、例えばイノシン、クエオシン (queosine)、ワイブトシン (wybutosine) を加える。

10

【0324】

CGDDをコードするポリヌクレオチドの発現の改変に有効な化合物をスクリーニングする方法が、本発明の更なる実施態様に含まれる。限定するものではないが特異ポリヌクレオチドの発現改変を起こすのに有効であり得る化合物には、オリゴヌクレオチド、アンチセンスオリゴヌクレオチド、三重らせん形成オリゴヌクレオチドや、転写因子などのポリペプチド転写制御因子、および、特異ポリヌクレオチド配列と相互作用し得る非高分子化学的実体がある。有効な化合物は、ポリヌクレオチド発現のインヒビターまたはエンハンサーのいずれかとして作用することによりポリヌクレオチド発現を改変し得る。従って、CGDDの発現または活性の増加に関連する疾患の治療においてはCGDDをコードするポリヌクレオチドの発現を特異的に阻害する化合物が治療上有用であり、CGDDの発現または活性の低下に関連する疾患の治療においてはCGDDをコードするポリヌクレオチドの発現を特異的に促進する化合物が治療上有用であり得る。

20

【0325】

或る特定ポリヌクレオチドの発現を改変する際の有効性について、少なくとも1個、または複数個の試験化合物をスクリーニングし得る。試験化合物を得るには、当分野で公知の任意の方法を用い得る。取得方法としては、以下の場合に有効な既知化合物の化学修飾がある。ポリヌクレオチドの発現を改変する場合と、既存の、市販のまたは私的な、天然または非天然の化合物のライブラリから選択する場合と、標的ポリヌクレオチドの化学的および／または構造的特性に基づく化合物を合理的にデザインする場合と、組合せ的にまたは無作為に生成した化合物のライブラリから選択する場合とである。CGDDをコードするポリヌクレオチドを持つサンプルを、このようにして得た試験化合物の少なくとも1つに曝露する。サンプルは例えば、無傷細胞、透過化処理した細胞、*in vitro*無細胞系すなわち再構成生化学系があり得る。CGDDをコードするポリヌクレオチドの発現の改変は、当分野で公知の任意の方法でアッセイする。通常、或る特定ヌクレオチドの発現を検出するにはCGDDをコードするポリヌクレオチドの配列に相補的なヌクレオチド配列を持つプローブとのハイブリダイゼーションを用いる。ハイブリダイゼーション量を定量することにより、1つ以上の試験化合物に曝露される、または曝露されないポリヌクレオチドの発現の比較のための基礎を形成し得る。試験化合物に曝露されるポリヌクレオチドの発現における変化の検出は、ポリヌクレオチドの発現を改変する際に試験化合物が有効であることを示す。或る特定ポリヌクレオチドの改変発現に有効な化合物のためのスクリーニングを実行でき、例えば分裂酵母 (*Schizosaccharomyces pombe*) 遺伝子発現系 (Atkins, D. 他(1999)

30

40

米国特許第5,932,435号、Arndt, G.M. 他(2000) *Nucleic Acids Res.*28:E15) またはHeLa細胞等のヒト細胞株 (Clarke, M.L. 他(2000) *Biochem. Biophys. Res. Commun.* 268:8-13) を用いる。本発明の或る特定の実施態様は、或る特定ポリヌクレオチド配列に対するアンチセンス活性について、オリゴヌクレオチド (デオキシリボヌクレオチド、リボヌクレオチド、ペプチド核酸、修飾したオリゴヌクレオチド) の組合せライブラリをスクリーニングする過程に関する (Bruice, T.W. 他(1997) 米国特許第5,686,242号、Bruice, T.W

50

.他(2000) 米国特許第6,022,691号)。

【0326】

ベクターを細胞または組織に導入する多数の方法が利用可能であり、in vivo、in vitro および ex vivo の使用に対して同程度に適している。ex vivo 治療の場合、ベクターを、患者から採取した幹細胞内に導入し、クローニング増殖して同患者に自家移植で戻すことができる。トランスフェクションによる、またはリボソーム注入やポリカチオンアミノポリマーによる送達は、当分野でよく知られている方法を用いて実行することができる(例えば Goldman, C.K. 他(1997) Nat. Biotechnol. 15:462-466を参照)。

【0327】

上記の治療方法はいずれも、例えば、ヒト、イヌ、ネコ、ウシ、ウマ、ウサギ、サルなどの哺乳類を含めて治療が必要な全ての被験体に適用できる。 10

【0328】

本発明の或る更なる実施態様は、薬物として許容できる或る賦形剤と共に製剤される或る活性成分を一般に有する、或る組成物の投与に関する。賦形剤には例えば、糖、でんぷん、セルロース、ゴムおよびタンパク質がある。様々な剤型が広く知られており、詳細は Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA) の最新版に記載されている。CGDD、CGDDの抗体、擬態物質、アゴニスト、アンタゴニスト、またはインヒビターなどが、このような組成物を構成しうる。

【0329】

本発明に用いる組成物は、任意の数の経路によって投与することができ、限定するものではないが経路には、経口、静脈内、筋肉内、動脈内、骨髄内、クモ膜下腔内、心室内、肺、経皮、皮下、腹腔内、鼻腔内、腸内、局所、舌下または直腸がある。 20

【0330】

肺から投与する組成物は、液状または乾燥粉末状で調製し得る。このような組成物は通常、患者が吸入する直前にエアロゾル化する。小分子(例えば伝統的な低分子量有機薬)の場合には、速効製剤のエアロゾル送達は当分野で公知である。高分子(例えばより大きなペプチドおよびタンパク質)の場合には、当該分野において肺の肺泡領域を介しての肺送達が最近向上したことにより、インスリンなどの薬物を実質的に血液循環へ輸送することを可能にした(Patton, J.S. 他, 米国特許第5,997,848号などを参照)。肺送達は、針注射なしに投与する点で優れており、有毒な可能性のある浸透エンハンサーの必要性をなく 30

【0331】

本発明での使用に適した組成物には、所定の目的を達成するために必要なだけの量の活性成分を含有する組成物が含まれる。有効投与量の決定は、当業者の能力の範囲内で行う。

【0332】

CGDDを有する、またはその断片群を有する高分子群を直接、細胞内に送達すべく、特殊な種々の形状の組成物群が調製され得る。例えば、細胞不透過性高分子を含むリポソーム製剤は、その高分子の細胞融合と細胞内送達とを促進し得る。または、CGDDまたはその断片を、HIV Tat-1タンパク質の陽イオンN末端部に結合することもできる。このようにして生成された融合タンパク質類は、或るマウスモデル系の、脳を含む全ての組織の細胞群に 40

。

【0333】

任意の化合物に対して、先ず細胞培養アッセイ、例えば新生物性細胞の細胞培養アッセイにおいて、あるいは、動物モデル、例えばマウス、ラット、ウサギ、イヌ、サルまたはブタなどにおいて、治療有効量を推定することができる。動物モデルはまた、好適な濃度範囲および投与経路を決定するためにも用い得る。このような情報を用いて、次にヒトに対する有益な投与量および投与経路を決定することができる。

【0334】

治療有効量は例えばCGDDやその断片、CGDDの抗体、アゴニストまたはアンタゴニスト、イ 50

ンヒビターなど、症状や容態を回復させる活性処方成分の量を指す。治療有効性および毒性は、細胞培養または動物実験における標準的な薬学手法によって、例えばED₅₀（集団の50%の治療有効量）またはLD₅₀（集団の50%の致死量）統計を計算するなどして判定できる。毒性効果の、治療効果に対する用量比が治療指数であり、LD₅₀/ED₅₀比として表すことができる。高い治療指数を示すような組成物が望ましい。細胞培養アッセイと動物実験とから得られたデータは、ヒトに用いる投与量の範囲の策定に用いられる。このような組成物に含まれる投与量は、毒性を殆どあるいは全く持たず、ED₅₀を含むような血中濃度の範囲にあることが好ましい。用いられる投与形態、患者の感受性および投与の経路によって、投与量はこの範囲内で様々に変わる。

【0335】

正確な投与量は、治療が必要な被験者に関する要素を考慮して、現場の医者が決定することになる。十分なレベルの活性成分を与え、あるいは所望の効果を維持すべく、用法および用量を調整する。被験者に関する要素としては、疾患の重症度、患者の一般的な健康状態、患者の年齢、体重および性別（ジェンダー）、投与の時間および頻度、薬物の配合、反応感受性および治療に対する応答などを考慮し得る。作用期間が長い組成物は、特定の製剤の半減期およびクリアランス率によって3～4日毎に1度、1週間に1度、あるいは2週間に1度の間隔で投与し得る。

【0336】

通常の投与量は、投与の経路にもよるが約0.1～100,000 μgであり、合計で約1gまでとする。特定の投与量および送達方法に関するガイダンスは文献に記載されており、現場の医者は通常それを利用することができる。当業者は、ヌクレオチドの処方では、タンパク質またはそれらのインヒビター類とは異なる処方を利用することになる。同様に、ポリヌクレオチドまたはポリペプチドの送達は、特定の細胞、症状、部位などに特異的なものとなる。

【0337】

（診断）

別の実施態様では、CGDDに特異的に結合する抗体を、CGDDの発現によって特徴付けられる障害の診断に、または、CGDDやそのアゴニストまたはアンタゴニスト、インヒビターで治療を受けている患者をモニターするためのアッセイに用い得る。診断目的に有用な抗体は、上記の治療の箇所で記載した方法と同じ方法で調合される。CGDDの診断アッセイには、抗体及び標識を用いて、ヒトの体液において、或いは細胞や組織の抽出物において、CGDDを検出する方法が含まれる。この抗体は、修飾して、または修飾せずに使用される。また、レポーター分子との共有結合または非共有結合で標識化できる。レポーター分子としては広くさまざまな種類が本分野で知られており、また使用可能であるが、そのうちのいくつかは上記で説明されている。

【0338】

CGDDを測定するための様々なプロトコル、例えばELISA、RIA、FACS等が当分野では周知であり、変容した、あるいは異常なレベルのCGDDの発現を診断するための基盤を提供する。正常或いは標準的CGDD発現の値は、複合体の形成に適した条件下で、正常な哺乳動物、例えばヒトなどの被験体から採取した体液または細胞抽出物とCGDDに対する抗体とを混合することによって確定する。標準複合体形成量は、種々の方法、例えば測光法で定量できる。CGDDの、被験体、対照および、生検組織からの疾患サンプルでの発現量を、標準値と比較する。標準値と被験体との偏差が、疾患を診断するパラメータを確定する。

【0339】

CGDDをコードするポリヌクレオチドを、本発明の別の実施態様では、診断目的で用い得る。用い得るポリヌクレオチドには、オリゴヌクレオチド配列、相補的RNAおよびDNA分子、そしてPNAが含まれる。CGDDの発現が疾患と関連し得る生検組織における遺伝子発現を、これらのポリヌクレオチドを用いて検出し定量し得る。CGDDの存在の有無、更には過剰な発現をこの診断アッセイを用いて判定し、治療時のCGDDレベルの調節を監視する。

【0340】

10

20

30

40

50

CGDDをコードまたは近縁の分子をコードするポリヌクレオチド配列（例えばゲノム配列）を検出可能なPCRプローブ類とのハイブリダイゼーションによって一態様では、CGDDをコードする核酸配列を同定できる。CGDDをコードする天然配列のみをプローブが同定するか、または対立遺伝子変異体や関連配列を同定するかは、プローブが高度に特異的な領域（例えば5'調節領域）から作られているか、またはやや特異性の低い領域（例えば保存されたモチーフ）から作られているかという、そのプローブの特異性と、ハイブリダイゼーションのまたは増幅のストリンジェンシーとによって決まることになる。

【0341】

CGDDをコードする任意の配列との少なくとも50%の配列同一性をもプローブ類は持ち得る他、関連する配列群の検出に利用され得る。本発明のハイブリダイゼーションプローブには、DNAあるいはRNAが可能であり、SEQ ID NO:22-42の配列や、CGDD遺伝子のプロモーター、エンハンサー、イントロンを含むゲノム配列に由来し得る。 10

【0342】

CGDDをコードするDNA群に対して特異的なハイブリダイゼーションプローブ群の作製方法としては、CGDDをコードまたはその誘導体群をコードするポリヌクレオチド配列群を、mRNAプローブ群の作製のためのベクター類にクローニングする方法を含む。mRNAプローブ作製用ベクター類は、当業者に知られており、市販されており、好適なRNAポリメラーゼと好適な標識されたヌクレオチドとを加えることによって、in vitroでRNAプローブを合成するために用い得る。ハイブリダイゼーションプローブ類は、種々のレポーターの集団によって標識され得る。レポーター集団の例としては、³²Pまたは³⁵Sなどの放射性核種が、あるいはアビジン／ビオチン結合系を介してプローブに結合されたアルカリホスファターゼなど酵素標識が挙げられる。 20

【0343】

CGDDをコードするポリヌクレオチド配列を、CGDDの発現に関係する疾患の診断の為に用い得る。限定するものではないが、このような疾患には細胞増殖異常、発達障害、神経疾患、自己免疫／炎症疾患、生殖障害、胎盤障害が含まれ、細胞増殖異常には日光角化症、動脈硬化、アテローム硬化、滑液包炎、硬変、肝炎、混合性結合組織病（MCTD）、骨髄線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症、並びに、腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫、および奇形癌など癌、具体的には、副腎、膀胱、骨、骨髄、脳、乳房、頸部、胆嚢、神経節、消化管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、および子宮の癌が含まれる。発達障害には尿細管性アシドーシス、貧血、クッシング症候群、軟骨形成不全性小人症、デュシェンヌ／ベッカー型筋ジストロフィー、癲癇、性腺形成異常、WAGR症候群（ウィルムス腫瘍、無虹彩症、尿生殖器異常、精神遅滞）、スミス－マジェニス症候群（Smith-Magenis syndrome）、骨髄異形成症候群、遺伝性粘膜上皮異形成、遺伝性角皮症や、シャルコーマリーツース病および神経線維腫症などの遺伝性神経病、甲状腺機能低下症、水頭症や、Syndenham舞蹈病（Syndenham's chorea）および脳性小児麻痺などの発作障害、二分脊椎、無脳症、頭蓋脊椎披裂、先天性緑内障、白内障、感音難聴が含まれる。神経疾患には、癲癇、虚血性脳血管障害、脳卒中、脳腫瘍、アルツハイマー病、ピック病、ハンチントン病、痴呆、パーキンソン病およびその他の錐体外路障害、筋萎縮性側索硬化およびその他の運動ニューロン障害、進行性神経性筋萎縮症、網膜色素変性症（色素性網膜炎）、遺伝性運動失調、多発性硬化症および他の脱髄疾患、細菌性およびウイルス性髄膜炎、脳膿瘍、硬膜下膿瘍、硬膜外膿瘍、化膿性頭蓋内血栓性静脈炎、脊髄炎および神経根炎、ウイルス性中枢神経系疾患と、プリオン病（クールー、クロイツフェルト－ヤコブ病、およびGerstmann-Straussler-Scheinker症候群を含む）、致死性家族性不眠症、神経系性栄養病および代謝病、神経線維腫症、結節硬化症、小脳網膜血管腫症（cerebelloretinal hemangioblastomatosis）、脳3叉神経血管症候群、ダウン症を含む中枢神経系性精神遅滞および他の発達障害、脳性麻痺、神経骨格異常症、自律神経系障害、脳神経疾患、脊髄疾患、筋ジストロフィー他の神経筋障害、末梢神経疾患、皮膚筋炎および多発性筋炎と、遺伝性、代謝性、内分泌性、および中毒性ミオパシーと、重症筋 40 50

無力症、周期性四肢麻痺、精神障害（気分性、不安性の障害、統合失調症／分裂病）、季節性感情障害（SAD）、静座不能、健忘症、緊張病、糖尿病性ニューロパシー、遅発性ジスキネジア、ジストニー、パラノイド精神病、帯状疱疹後神経痛、トゥーレット病、進行性核上麻痺、大脳皮質基底核変性（corticobasal degeneration）、および家族性前頭側頭型痴呆とが含まれる。自己免疫／炎症疾患には、後天性免疫不全症候群（AIDS）、アジソン病（慢性原発性副腎機能不全）、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫性多腺性内分泌カンジダ症外胚葉ジストロフィー（APECED）、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、糖尿病、肺気腫、リンパ球傷害因子性偶発性リンパ球減少症、新生児溶血性疾患（胎児赤芽球症）、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性腸症候群、多発性硬化症、重症筋無力症、心筋または心膜炎、骨関節炎、骨粗しょう症、腓炎、多発性筋炎、乾癬、ライター症候群、リウマチ様関節炎、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性強皮症、血小板減少性紫斑病、潰瘍性大腸炎、ぶどう膜炎、ウェルナー症候群、癌合併症、血液透析、体外循環、ウイルス感染症、細菌感染症、真菌感染症、寄生虫感染症、原虫感染症、蠕虫感染症、および外傷が含まれる。生殖障害には、プロラクチン産生の障害、不妊（例えば卵管疾患、排卵不良、子宮内膜症、性周期の途絶、月経周期の途絶、多嚢胞性卵巣症候群、卵巣過剰刺激症候群、子宮内膜腫瘍や卵巣腫瘍、子宮筋腫、自己免疫疾患、異所性妊娠、奇形発生、乳癌、乳房線維嚢胞病、乳漏症、精子形成の途絶、異常精子生理、精巣癌、前立腺癌、良性前立腺肥大、前立腺炎、ペーロニー病、インポテンス、男性乳癌、女性化乳房、高ゴナドトロピン性腺機能低下症、低ゴナドトロピン性腺機能低下症、仮性半陰陽、無精子症、早発卵巣不全、アクロシン欠損症、晩発思春期、逆行性射精、無射精、血管芽腫、嚢胞・クロム親和細胞腫（cystsphaeochromocytomas）、傍神経節腫、精巣上体の嚢胞腺腫、および内リンパ嚢腫瘍）を含む。胎盤障害には、子癇前症、絨毛癌、胎盤早期剥離、前置胎盤、胎盤梗塞やmaternal floor infarction、癒着胎盤、侵入胎盤（increate）および穿通胎盤、絨毛膜外性胎盤（extrachorial placentas）、絨毛血管腫（chorangioma）、絨毛血管分布過多（chorangiosis）、慢性絨毛炎（villitis）、胎盤絨毛浮腫（placental villous edema）、終末絨毛広範性線維症（widespread fibrosis of the terminal villi）、絨毛間腔血栓（intervillous thrombi）、出血性血管炎、新生児溶血性疾患（胎児赤芽球症）、非免疫性胎児水腫を含む。CGDDをコードするポリヌクレオチド配列は、変容したCGDD発現を検出するために患者から採取した体液或いは組織を利用するサザン法やノーザン法、ドットプロット法、或いはその他の膜系の技術、PCR法、ディップスティック（dipstick）、ピン（pin）、およびマルチフォーマットのELISA式アッセイ、並びにマイクロアレイに使用可能である。このような定性方法または定量方法は、当分野で公知である。

10

20

30

【0344】

CGDDをコードするヌクレオチド配列群は、或る特定の態様では、関連する障害、特に上記した障害を検出するアッセイ類において有用であり得る。CGDDをコードするヌクレオチド配列は、標準的な方法で標識化され、ハイブリダイゼーション複合体の形成に好適な条件下で、患者から採取した体液あるいは組織のサンプルに加えることができる。好適なインキュベーション期間が経過したらサンプルを洗浄し、シグナルを定量して標準値と比較する。患者のサンプルのシグナルの量が対照サンプルと比較して著しく変化している場合は、そのサンプル内のCGDDをコードするヌクレオチド配列群のレベル変化の存在により、関連する疾患の存在が明らかになる。このようなアッセイは、動物実験、臨床試験における特定の治療効果を評価するため、あるいは個々の患者の治療をモニターするために用いることもできる。

40

【0345】

CGDDの発現に関連する疾患の診断の基準を提供するために、発現の正常すなわち標準的なプロファイルが確立される。これは、ハイブリダイゼーションあるいは増幅に好適な条件

50

の下、動物あるいはヒトのいずれかの正常な被験体から抽出された体液あるいは細胞とCGDDをコードする配列あるいはその断片とを混合することにより達成され得る。実質的に精製されたポリヌクレオチドを既知量で用いて行った実験から得た値を正常な被験者から得た値と比較することにより、標準ハイブリダイゼーションを定量することができる。このようにして得た標準値は、疾患の徴候を示す患者から得たサンプルから得た値と比較することができる。標準値からの偏差を用いて疾患の存在を確定する。

【0346】

疾患の存在が確定されて治療プロトコルが開始されると、患者の発現レベルが正常な被検者に観察されるレベルに近づき始めたかどうかを測定するため、ハイブリダイゼーションアッセイを定期的に繰り返し得る。連続アッセイから得られた結果を用いて、数日から数ヶ月の期間にわたる治療の効果を示し得る。 10

【0347】

癌に関しては、個体からの生体組織における異常な量の転写物（過少発現または過剰発現）の存在は、疾患の発生素質を示したり、実際に臨床的症状が現れる前に疾患を検出する方法を提供し得る。この種のより明確な診断により、医療の専門家が予防方法または積極的な治療法を早くから利用し、それによって癌の発生または更なる進行を防止することが可能となる。

【0348】

CGDDをコードする配列群から設計したオリゴヌクレオチド群の更なる診断的利用には、PCRの利用を含み得る。これらのオリゴマーは、化学的に合成するか、酵素により生産するか、あるいは*in vitro*で産出し得る。オリゴマーはCGDDをコードするポリヌクレオチドの断片や、CGDDをコードするポリヌクレオチドと相補的なポリヌクレオチドの断片を好適には含み、最適な条件下で、特定の遺伝子や条件を識別するために利用される。また、オリゴマーは、やや緩いストリンジェンシー条件下で、近縁のDNAあるいはRNA配列の検出、定量、あるいはその両方のため用いることが可能である。 20

【0349】

CGDDをコードするポリヌクレオチド配列群に由来するオリゴヌクレオチドプライマー類を用いて、或る特定態様においては、一塩基多型（SNP）を検出し得る。SNPは、多くの場合にヒトの先天性または後天性遺伝病の原因となるような置換、挿入および欠失である。限定するものではないがSNPの検出方法には、SSCP（single-stranded conformation polymorphism）および蛍光SSCP（fSSCP）がある。SSCPはCGDDをコードするポリヌクレオチド配列に由来するオリゴヌクレオチドプライマーを用いたポリメラーゼ連鎖反応（PCR）でDNAを増幅する。このDNAは例えば、病変組織または正常組織、生検サンプル、体液などに由来し得る。DNA内のSNPは、一本鎖形のPCR生成物の2次および3次構造に差異を生じさせる。差異は非変性ゲル中でのゲル電気泳動法を用いて検出可能である。fSSCPでは、オリゴヌクレオチドプライマーを蛍光性に標識する。それによってDNAシーケンシング機などの高処理機器でアンプリマー（amplimer）の検出が可能になる。更に、インシリコSNP（*in silico* SNP, isSNP）と呼ばれる配列データベース分析法は、共通のコンセンサス配列へアセンブリされるような個々のオーバーラップするDNA断片群の配列を比較することにより、多型性を同定し得る。これらのコンピュータベースの方法は、DNAの実験室での調製に、または統計モデルとDNA配列クロマトグラムの自動分析とを用いたシーケンシングのエラーに起因する配列変異を、フィルタリングして除去する。別の態様では、例えば高処理MASSARRAYシステム（Sequenom, Inc., San Diego CA）を用いた質量分析によりSNPを検出し、特徴付ける。 30 40

【0350】

SNPを利用して、ヒト疾患の遺伝的基礎を研究しうる。例えば少なくとも16種の一般的SNPが、インスリン非依存型糖尿病と関連している。SNPはまた、単一遺伝子病における疾病転帰における相違を試験するにも有用である。単一遺伝子病としては、嚢胞性線維症、鎌状赤血球貧血、または慢性肉芽腫症がある。例えばマンノース結合レクチンであるMBL2における変異体群は、嚢胞性線維症における有害な肺の転帰との相関が示されている。SNP 50

はまた、薬理ゲノム学においても有用性を持つ。薬理ゲノム学は、患者の薬物への応答（例えば命を脅かす毒性）に影響する遺伝的変異体の同定を行う。例えばNアセチル転移酵素における或る変異は、抗結核薬であるイソニアジドに反応しての末梢ニューロパシーの高い発症率と関連する。一方、ALOX5遺伝子のコアプロモータにおける或る変異の結果、5-リポキシゲナーゼ経路を標的とする或る抗喘息薬での治療に対する臨床応答が低下する。異なる集団群におけるSNPの分布の分析は、遺伝的浮動、突然変異、遺伝子組換え、遺伝子選択の調査に有用であり、集団の起源とその移動との追跡にも有用である(Taylor, J.G. 他(2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. および Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. 他 (2001) Curr. Opin. Neurobiol. 11:637-641)。

【0351】

CGDDの発現を定量するために用い得る別の方法の例としては、ヌクレオチド群の放射標識またはビオチン標識、対照核酸の共増幅(coamplification)、および、標準曲線から得た結果の補間もある(例えばMelby, P.C.他(1993) J. Immunol. Methods 159:235-244; Duplaa, C.他(1993) Anal. Biochem. 212:229-236を参照)。複数サンプルの定量速度を加速するには、目的のオリゴマーまたはポリヌクレオチドを種々の希釈液中に置き、分光光度法または比色反応によって定量が迅速になるような高処理フォーマットでのアッセイを行い得る。

【0352】

更に別の実施例では、本明細書で記載した任意のポリヌクレオチド配列に由来するオリゴヌクレオチドまたはより長い断片を、或るマイクロアレイにおけるエレメント群として用いることができる。多数の遺伝子の相対発現レベルを同時にモニターする転写イメージング技術にマイクロアレイを用いることが可能である。これについては、以下に記載する。マイクロアレイはまた、遺伝変異体、突然変異および多型性の同定に用いることができる。この情報を用いることで、遺伝子機能を決定し、疾患の遺伝的根拠を理解し、疾患を診断し、遺伝子発現の機能としての疾病の進行/後退をモニターし、疾病治療における薬物の活性を開発およびモニターすることができる。特に、患者にとって最もふさわしく、有効な治療法を選択するために、この情報を用いて患者の薬理ゲノムプロファイルを開発することができる。例えば、患者の薬理ゲノムプロファイルに基づき、患者に対して高度に効果的で副作用を殆ど示さない治療薬を選択することができる。

【0353】

CGDDやその断片群、CGDDに特異的な抗体類を、別の実施態様では、或るマイクロアレイ上のエレメント群として用い得る。マイクロアレイを用いて、上記のようなタンパク質-タンパク質相互作用、薬物-標的相互作用および遺伝子発現プロファイルをモニターまたは測定することが可能である。

【0354】

或る実施態様は、或る組織または細胞タイプの転写イメージを作製する、本発明のポリヌクレオチドの使用に関連する。転写イメージは、特定の組織または細胞タイプによる、遺伝子発現の包括的パターンを表す。包括的遺伝子発現パターンは、所与の条件下で所与の時間に発現した遺伝子の数および相対存在量を定量することにより分析し得る(Seilhame r 他(米国特許第5,840,484号「Comparative Gene Transcript Analysis」を参照。該特許は特に引用することを以って本明細書の一部となす)。したがって、特定の組織または細胞タイプの転写物または逆転写物全体に、本発明のポリヌクレオチドまたはそれらの相補体をハイブリダイズすることにより、転写イメージを作製し得る。或る実施態様では、本発明のポリヌクレオチドまたはそれらの相補体が1マイクロアレイ上に複数エレメントの1サブセットを持つような高処理フォーマットでハイブリダイゼーションを発生させる。結果として得られる転写イメージは、遺伝子活性のプロファイルを提供し得る。

【0355】

転写イメージは、組織、細胞株、生検またはその生体サンプルから単離した転写物を用いて作製し得る。転写イメージはしたがって、組織または生検サンプルの場合にはin vivo、細胞株の場合にはin vitroでの遺伝子発現を反映する。

10

20

30

40

50

【0356】

本発明のポリヌクレオチドの発現のプロファイルを作製する転写イメージはまた、in vitroモデル系および薬物の前臨床評価に、あるいは工業的または天然の環境化合物の毒性試験に関連して使用し得る。全ての化合物は、作用および毒性のメカニズムを標示し、しばしば分子フィンガープリントまたは毒性シグネチャ(toxicant signatures)と称される、特徴的な遺伝子発現パターンを惹起する(Nuwaysir, E.F. 他(1999) Mol. Carcinog. 24:153-159、Steiner, S. および N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471、該文献は特に引用を以って本明細書の一部となす)。試験化合物が、既知毒性を有する化合物のシグネチャと同一のシグネチャを有する場合には、毒性特性を共有している可能性がある。フィンガープリントまたはシグネチャは、多数の遺伝子および遺伝子ファミリーからの発現情報を含んでいる場合に、最も有用且つ正確である。理想的には、ゲノム全域にわたる発現の測定が、最高品質のシグネチャを提供する。たとえ、発現が任意の試験された化合物によって変容しない遺伝子があったとしても、それらの発現レベルを残りの発現データをノーマライズすることによって使用できるため、それらの遺伝子は重要である。ノーマライズ手順は、種々の化合物で処理した後の発現データの比較に有用である。或る毒性シグネチャのエレメント群に遺伝子機能を割り当てることは毒性機序の解釈に役立つが、毒性の予測につながる、シグネチャ群の統計的マッチングには、遺伝子機能の知識は必要でない(例えば2000年2月29日に米国国立環境健康科学研究所(National Institute of Environmental Health Sciences)より発行されたPress Release 00-02を参照されたい。これについては<http://www.niehs.nih.gov/oc/news/toxchip.htm>で入手可能である)。したがって、毒性シグネチャを用いる中毒学的スクリーニングの際に、全ての発現した遺伝子配列を含めることは、重要且つ望ましい。

10

20

【0357】

或る実施例では、核酸を有する生体サンプルを試験化合物で処理することにより、この試験化合物の毒性を算定する。処理した生体サンプル中で発現した核酸は、本発明のポリヌクレオチドに特異的な1つ以上のプローブでハイブリダイズし、それによって本発明のポリヌクレオチドに対応する転写レベルを定量し得る。処理した生体サンプル中の転写レベルを、未処理生体サンプル中のレベルと比較する。両サンプルの転写レベルの差は、処理済サンプル中で試験化合物が引き起こす毒性反応を標示する。

【0358】

別の実施態様は、本発明のポリペプチド配列群を用いて或る組織または細胞タイプのプロテオームを分析することに関する。プロテオームの語は、特定の組織または細胞タイプでのタンパク質発現の包括的パターンを指す。プロテオームの各タンパク質成分は、個々に更なる分析の対象とすることができる。プロテオーム発現パターンすなわちプロファイルは、所与の条件下で所与の時間に発現したタンパク質の数および相対存在量を定量することにより分析し得る。したがって、或る細胞のプロテオームのプロファイルは、特定の組織または細胞タイプのポリペプチドを分離および分析することにより作成し得る。或る実施例では、1次元等電点電気泳動によりサンプルからタンパク質を分離し、2次元ドデシル硫酸ナトリウムスラブゲル電気泳動により分子量に応じて分離するような2次元ゲル電気泳動により、分離が達成される(前出のSteiner および Anderson)。タンパク質は、通常はクーマシーブルー、あるいは銀染色液または蛍光染色液などの物質を用いてゲルを染色することにより、分散した、独自の位置にある点としてゲル中で可視化される。各タンパク質スポットの光学密度は、通常、サンプル中のタンパク質レベルに比例する。異なるサンプル、例えば試験化合物または治療薬で処理または未処理のいずれかの生体サンプルからの、同等に位置したタンパク質スポットの光学密度を比較し、処理に関連するタンパク質スポット密度の変化を同定する。スポット内のタンパク質は、例えば化学的または酵素的切断とそれに続く質量分析を用いる標準的な方法を用いて部分的にシークエンシングする。或るスポット内のタンパク質の同一性は、その部分配列を、好適には少なくとも5個の連続するアミノ酸残基を、本発明のポリペプチド配列と比較することにより決定し得る。場合によっては、決定的なタンパク質同定のための更なる配列データが得られる。

30

40

50

【0359】

CGDD特異抗体でCGDD発現レベルを定量することによってもプロテオームのプロファイルを作成し得る。或る実施態様では、マイクロアレイ上のエレメントとしてこれら抗体を用い、マイクロアレイをサンプルに曝して各アレイエレメントへのタンパク質結合レベルを検出することにより、タンパク質発現レベルを定量する (Lueking, A. 他(1999) Anal. Biochem. 270:103-111; Mendoz, L.G.他(1999) Biotechniques 27:778-788)。検出は当分野で既知の様々な方法で行うことができ、例えば、チオール反応性またはアミノ反応性蛍光化合物とサンプル中のタンパク質を反応させ、各アレイエレメントにおける蛍光結合の量を検出し得る。

【0360】

プロテオームレベルでの毒性シグネチャも中毒学的スクリーニングに有用であり、転写レベルでの毒性シグネチャと並行に分析するべきである。いくつかの組織のいくつかのタンパク質については、転写物の存在量とタンパク質の存在量との相関が乏しいので (Anderson, N.L. および J. Seilhamer (1997) Electrophoresis 18:533-537)、転写イメージにはそれ程影響しないがプロテオームのプロファイルを改変するような化合物の分析において、プロテオーム毒性シグネチャは有用たり得る。更に、体液中の転写物の分析はmRNAの急速な分解のために困難なので、プロテオームのプロファイル作成はこのような場合により信頼し得、情報価値があり得る。

10

【0361】

別の実施例では、タンパク質を含有する生体サンプルを試験化合物で処理することにより試験化合物の毒性を算定する。処理された生体サンプル中で発現したタンパク質は、各タンパク質の量を定量し得るように分離する。各タンパク質の量を、未処理生体サンプル中の対応するタンパク質の量と比較する。両サンプルのタンパク質の量の差は、処理サンプル中の試験化合物に対する毒性反応を標示する。個々のタンパク質は、個々のタンパク質のアミノ酸残基をシーケンシングし、これら部分配列を本発明のポリペプチドと比較することにより同定する。

20

【0362】

別の実施例では、タンパク質を含有する生体サンプルを試験化合物で処理することにより試験化合物の毒性を算定する。生体サンプルから得たタンパク質は、本発明のポリペプチドに特異的な抗体を用いてインキュベートする。抗体により認識されたタンパク質の量を定量する。処理された生体サンプル中のタンパク質の量を、未処理生体サンプル中のタンパク質の量と比較する。両サンプルのタンパク質の量の差は、処理サンプル中の試験化合物に対する毒性反応を標示する。

30

【0363】

マイクロアレイは、本技術分野で既知の方法で調製し、使用し、分析する (例えばBrennan, T.M. 他 (1995) 米国特許第5,474,796号、Schena, M. 他 (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619、Baldeschweiler 他 (1995) PCT出願第W095/251116号、Shalon, D.他 (1995) PCT出願第W095/35505号、Heller, R.A. 他 (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155、Heller, M.J. 他 (1997) 米国特許第5,605,662号を参照)。様々なタイプのマイクロアレイが公知であり、詳細については、DNA Microarrays: A Practical Approach, M. Schena, 編集 (1999) Oxford University Press, Londonに記載がある。該文献は、特に引用を以って本明細書の一部となす。

40

【0364】

CGDDをコードする核酸配列を用いて、本発明の別の実施態様では、天然のゲノム配列をマッピングするのに有用なハイブリダイゼーションプローブを作製することが可能である。コード配列または非コード配列のいずれかを用いることができ、いくつかの例では、コード配列より非コード配列が好ましい。例えば、多重遺伝子族のメンバー内でのコード配列の保存により、染色体マッピング中に望ましくないクロスハイブリダイゼーションが生じる可能性がある。配列は、或る特定の染色体に、または或る染色体の或る特定領域に、または人為形成の染色体、例えば、ヒト人工染色体 (HAC)、酵母人工染色体 (YAC)、細菌

50

人工染色体（BAC）、細菌P1産物、あるいは単一染色体cDNAライブラリ群に対してマッピングされる（例えばHarrington, J.J.他（1997）Nat. Genet. 15:345-355; Price, C.M.（1993）Blood Rev. 7:127-134; Trask, B.J.（1991）Trends Genet. 7:149-154を参照）。一度マッピングすると、本発明の核酸配列群を用いて、例えば或る病状の遺伝を特定染色体領域の遺伝とまたは制限酵素断片長多型（RFLP）と相関させるような遺伝子連鎖地図を開発し得る（例えば、Lander, E.S.およびD. Botstein（1986）Proc. Natl. Acad. Sci. USA 83:7353-7357を参照）。

【0365】

蛍光原位置ハイブリッド形成法（FISH）は、他の物理的および遺伝的地図データと相関し得る（例えばHeinz-Ulrich,他（1995）in Meyers, 前出、965-968ページを参照）。遺伝地
10 図データの例は、種々の科学雑誌あるいはOnline Mendelian Inheritance in Man（OMIM）のウェブサイトに見られる。CGDDをコードする遺伝子の物理地図上の位置と、特定の疾患との相関性、あるいは特定の疾患に対する素因との相関性は、この疾患と関連するDNAの領域の決定に役立ち得るため、位置を決定するクローニングの作業を促進し得る。

【0366】

確定した染色体マーカー類を用いた連鎖分析などの物理的マッピング技術、および染色体標本の原位置ハイブリッド形成法を用いて、遺伝地図を拡張することができる。例えばマウスなど別の哺乳類の染色体上に遺伝子を配置することにより、正確な染色体上の遺伝子座が未知でも、関連するマーカー類をしばしば明らかにし得る。この情報は、位置クロー
20 ニングなどの遺伝子発見技術を用いて疾患遺伝子を探る研究者にとって価値がある。いったん疾患または症候群に関与する遺伝子（群）が、血管拡張性失調症の11q22-23領域など、特定のゲノム領域への遺伝的連鎖によって、大まかに位置決めがなされると、該領域にマップされる任意の配列は、更なる調査のための関連遺伝子あるいは調節遺伝子を提示している可能性がある（例えばGatti, R.A. 他（1988）Nature 336:577-580を参照）。転座、反転などに起因する、健常者、保有者、罹病者の三者間における染色体位置の相違を検出する場合にも、本発明のヌクレオチド配列を用い得る。

【0367】

CGDD、その触媒作用断片あるいは免疫原性断片またはそのオリゴペプチドを、本発明の別の実施態様では、種々の任意の薬剤スクリーニング技術における、化合物のライブラリ類のスクリーニングに用い得る。薬剤スクリーニングに用いる断片は、溶液中に遊離してい
30 るか、固体支持物に固定されるか、細胞表面上に保持されるか、細胞内に位置することができる。CGDDと試験する薬剤との結合による複合体の形成を測定し得る。

【0368】

別の薬剤スクリーニング方法は、目的のタンパク質に対して好適な結合親和性を有する化合物を高い処理能力でスクリーニングするために用いられる（例えばGeysen, 他（1984）PCT出願第W084/03564号を参照）。この方法においては、多数の様々な低分子の試験用化合物を固体基板上で合成する。CGDDやその断片と反応してから試験用化合物を洗浄する。結合CGDDを次に当分野で周知の方法で検出する。精製CGDDはまた、上記した薬剤スクリー
40 ニング技術に用いるプレート上に直接コーティングできる。別法では、非中和抗体を用いてペプチドを捕捉し、ペプチドを固体支持物に固定することもできる。

【0369】

CGDDと特異結合可能な中和抗体がCGDDとの結合について試験化合物と競合する、競合的薬剤スクリーニングアッセイを別の実施態様で用いる。CGDDと1つ以上の抗原決定基を共有するどのペプチドの存在をも、この方法で抗体を用いて検出できる。

【0370】

CGDDをコードするヌクレオチド配列を、別の実施例では、将来に開発される分子生物学技術であり、現在知られているヌクレオチド配列の特性（限定はされないが、トリプレット遺伝コード、特異的な塩基対相互作用等を含む）に依存する新技術に用い得る。

【0371】

更に詳細説明をしなくとも、当業者であれば以上の説明を以って本発明を最大限に利用で
50

きるであろう。したがって、これ以下に記載する実施例は単なる例示目的にすぎず、いかようにも本発明を限定するものではない。

【0372】

上述し後述する全ての特許、特許出願および刊行物の開示は、米国特許出願第60/286,820号、第60/293,727号、第60/283,294号、第60/282,110号、第60/287,228号、第60/291,846号、第60/291,662号、第60/295,340号、第60/295,263号、第60/349,705号を含め、言及することを以て特に本明細書の一部となす。

【実施例】

【0373】

1 cDNAライブラリの作製

Incyte cDNA群の由来は、LIFESEQ GOLDデータベース (Incyte Genomics, Palo Alto CA) に記載されたcDNAライブラリ群である。幾つかの組織はホモジナイズしてグアニジニウムイソチオシアネート溶液に溶解し、他の組織はホモジナイズしてフェノールにまたは変性剤群の好適な混合液に溶解した。混合液の1例であるTRIZOL (Life Technologies) は、フェノールとグアニジンイソチオシアネートとの単相溶液である。結果として得た溶解物は、塩化セシウムのクッション液の上に重層して遠心分離するか、クロロホルムで抽出した。イソプロパノールか、酢酸ナトリウムとエタノールか、いずれか一方、あるいは別の慣例的方法で、溶解物からRNAを沈殿させた。

【0374】

RNAの純度を高めるため、フェノールによるRNAの抽出および沈殿を、必要な回数繰り返した。場合によっては、DNアーゼでRNAを処理した。殆どのライブラリでは、オリゴd(T) 連結常磁性粒子 (Promega)、OLIGOTEXラテックス粒子 (QIAGEN, Chatsworth CA) またはOLIGOTEX mRNA精製キット (QIAGEN) を用いて、ポリ(A)+RNAを単離した。別法では、別のRNA単離キット、例えばPOLY(A) PURE mRNA精製キット (Ambion, Austin TX) を用いて、組織溶解物からRNAを直接単離した。

【0375】

場合によってはStratagene社にRNAを提供し、同社が、対応するcDNAライブラリ群を作製した。そうでない場合は、UNIZAPベクターシステム (Stratagene) またはSUPERSCRIPPTプラスミドシステム (Life Technologies) を用いて本技術分野で既知の推奨方法または類似の方法でcDNAを合成し、cDNAライブラリを作製した (前出のAusubel, 1997, 5.1-6.6ユニットなどを参照)。逆転写は、オリゴd(T) またはランダムプライマーを用いて開始した。合成オリゴヌクレオチドアダプターを二本鎖cDNAに連結反応させ、好適な制限酵素または酵素群でcDNAを消化した。殆どのライブラリに対して、cDNAのサイズ選択 (300~1000 bp) は、SEPHACRYL S1000、SEPHAROSE CL2BまたはSEPHAROSE CL4Bカラムクロマトグラフィ (Amersham Pharmacia Biotech)、あるいは分取用アガロースゲル電気泳動法を用いて行った。cDNAは好適なプラスミドのポリリンカーの、適合する制限酵素部位にライゲーションされた。好適なプラスミドは、例えばPBLUESCRIPTプラスミド (Stratagene)、PSPORT1プラスミド (Life Technologies) PCDNA2.1プラスミド (Invitrogen, Carlsbad CA)、PBK-CMVプラスミド (Stratagene)、PCR2-TOPOTAプラスミド (Invitrogen)、PCMV-ICISプラスミド (Stratagene)、pIGEN (Incyte Genomics, Palo Alto CA)、pRARE (Incyte Genomics)、またはpINCY (Incyte Genomics)、またはこれらの誘導体である。組換えプラスミドは、Stratagene社のXL1-Blue、XL1-BlueMRFまたはSOLR、あるいはLife Technologies社のDH5α、DH10BまたはElectroMAX DH10Bなど適格な大腸菌細胞に形質転換した。

【0376】

2 cDNAクローンの単離

実施例1のようにして得たプラスミドの、宿主細胞からの回収は、UNIZAPベクターシステム (Stratagene) を用いた *in vivo* 切除によって、あるいは細胞溶解によって行った。プラスミドを精製する方法は、MagicまたはWIZARD Minipreps DNA精製システム (Promega)、AGTC Miniprep精製キット (Edge Biosystems, Gaithersburg MD)、QIAGEN社のQIAWELL

10

20

30

40

50

8 Plasmid、QIAWELL 8 Plus PlasmidおよびQIAWELL 8 Ultra Plasmid 精製システム、R. E.A.L. Prep 96プラスミド精製キットの中から少なくとも1つを用いた。プラスミドは、沈殿させた後、0.1mlの蒸留水に再懸濁して、凍結乾燥して或いは凍結乾燥しないで4℃で保管した。

【0377】

別法では、高処理フォーマットで直接結合PCR法を用い、宿主細胞溶解物からプラスミドDNAを増幅した(Rao, V.B. (1994) Anal. Biochem. 216:1-14)。宿主細胞の溶解および熱サイクリング過程は、単一反応混合液中で行った。サンプルを加工し、それを384穴プレート内で保管し、増幅したプラスミドDNAの濃度をPICOGREEN色素(Molecular Probes, Eugene OR) およびFLUOROSKAN II蛍光スキャナ(Labsystems Oy, Helsinki, Finland)を用いて蛍光分析的に定量した。 10

【0378】

3 シークエンシングおよび分析

実施例2に記載したようにプラスミドから回収したIncyte cDNAを、以下に示すようにシークエンシングした。cDNAのシークエンス反応は、標準的方法あるいは高処理装置、例えばABI CATALYST 800 サーマルサイクラー(Applied Biosystems)またはPTC-200 サーマルサイクラー(MJ Research)を、HYDRAマイクロディスペンサー(Robbins Scientific)またはMICROLAB 2200(Hamilton)液体転移システムと併用して処理した。cDNAのシークエンス反応は、Amersham Pharmacia Biotech社が提供する試薬、またはABIシークエンシングキット、例えばABI PRISM BIGDYE Terminator cyclesequencing ready reaction kit (Applied Biosystems)の試薬を用いて準備した。cDNAのシークエンス反応の電気泳動的分離には、また、標識したポリヌクレオチドの検出には、MEGABACE 1000 DNAシークエンシングシステム(Molecular Dynamics)か、標準ABIプロトコルと塩基対呼び出しソフトウェアとを用いるABI PRISM 373または377シークエンシングシステム(Applied Biosystems)か、あるいはその他の本技術分野で既知の配列解析システムを用いた。cDNA配列内のリーディングフレームは、標準的方法(前出のAusubel, 1997, 7.7ユニットに概説)を用いて同定した。いくつかのcDNA配列を選択し、実施例8に開示する技術で配列を伸長させた。 20

【0379】

Incyte cDNAに由来するポリヌクレオチド配列は、ベクター、リンカーおよびポリ(A)配列を除去し、あいまいな塩基をマスクすることによって有効性を確認した。その際、BLASTと、動的プログラミングと、隣接ジヌクレオチド頻度分析とに基づく、アルゴリズムとプログラムとを用いた。次に、Incyte cDNA配列またはそれらの翻訳の問い合わせを、以下のデータベース群に対して行った。すなわち、選抜した公共のデータベース群(例えばGenBankの霊長類、げっ歯類、哺乳類、脊椎動物、真核生物のデータベースと、BLOCKS、PRINTS、DOMO、PRODOM)と、ヒト、ラット、マウス、線虫(Caenorhabditis elegans)、出芽酵母(Saccharomyces cerevisiae)、分裂酵母(Schizosaccharomyces pombe)およびCandida albicansからの配列群を持つPROTEOMEデータベース群(Incyte Genomics, Palo Alto CA)、および、隠れマルコフモデル(HMM)ベースのタンパク質ファミリーデータベース群、例えばPFAM、INCY、およびTIGRFAM(Haft, D.H. 他(2001) Nucleic Acids Res. 29:41-43); 並びにHMMベースのタンパク質ドメインデータベースたとえばSMART(Schultz 他(1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. 他(2002) Nucleic Acids Res. 30:242-244)(HMMは、遺伝子ファミリーのコンセンサス1次構造を分析する確率的アプローチである。例えばEddy, S.R. (1996) Cuff. Opin. Struct. Biol. 6:361-365等を参照)。問合せは、BLAST、FASTA、BLIMPS、およびHMMERに基づくプログラムを用いて行った。Incyte cDNA配列を構築し、完全長のポリヌクレオチド配列を産出した。あるいは、GenBank cDNA群、GenBank EST群、スティッチされた配列群、ストレッチされた配列群、またはGenscan予測コード配列群(実施例4および5を参照)を用い、Incyte cDNAの構築物を完全長まで伸長させた。PhredとPhrapとConsedとに基づくプログラムを用いて構築し、GenMarkとBLASTとFASTAとに基づくプログラムを用いて、cDNAの構築物を、オー 30 40 50

ブンリーディングフレームについてスクリーニングした。完全長ポリヌクレオチド配列を翻訳し、対応する完全長ポリペプチド配列を誘導した。あるいは、本発明のポリペプチドは、完全長翻訳ポリペプチドの任意のメチオニン残基で開始し得る。完全長ポリペプチド配列群の続いての分析としての問い合わせを、GenBankタンパク質データベース群 (genpept)、SwissProt、PROTEOMEデータベース群、BLOCKS、PRINTS、DOMO、PRODOMおよびProsite等のデータベースや、PFAM、INCY、およびTIGRFAM等の隠れマルコフモデル (HMM) ベースのタンパク質ファミリーデータベース群、並びにSMART等のHMMベースのタンパク質ドメインデータベース群に対し行った。完全長ポリヌクレオチド配列はまた、MACDNASIS PROソフトウェア (日立ソフトウェアエンジニアリング, South San Francisco CA) およびLA SERGENEソフトウェア (DNASTAR) を用いて分析する。ポリヌクレオチドとポリペプチドとの配列アラインメントは、MEGALIGNマルチシークエンスアラインメントプログラム (DNASTAR) に組み込まれたCLUSTALアルゴリズムが指定する、デフォルトパラメータを用いて作製する。これは、アラインメントした配列間の一致率も計算する。

10

【0380】

表7は、Incyte cDNAと完全長配列との分析と構築とに利用したツールとプログラムとアルゴリズムとの概略と、適用可能な説明、参照文献、閾値パラメータを示す。用いたツール、プログラムおよびアルゴリズムを表7の列1に、それらの簡単な説明を列2に示す。列3は好適な参照文献であり、全ての文献は引用を以って全文を本明細書の一部となす。適用可能な場合には、列4は、2つの配列が一致する強さを評価するために用いた、スコア、確率値などのパラメータを示す (スコアが高いほど、または確率値が低いほど、2配列間の同一性が高い)。

20

【0381】

完全長ポリヌクレオチド配列群とポリペプチド配列群との構築と分析とに用いた上記プログラム群は、SEQ ID NO:22-42のポリヌクレオチド配列断片群の同定にも利用した。ハイブリダイゼーション技術と増幅技術とに有用な約20～約4000ヌクレオチドの断片群を、表4の列2に示した。

【0382】

4 ゲノムDNAからのコード配列の同定および編集

推定の細胞成長、分化および細胞死関連タンパク質は、先ず公共のゲノム配列データベース (例えばgbpriやgbhtg) に対しGenscan遺伝子同定プログラムを実行して同定した。Genscanは汎用遺伝子同定プログラムであり、様々な生物からのゲノムDNA配列を分析する (Burge, C. および S. Karlin (1997) J. Mol. Biol. 268:78-94、Burge, C. および S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354参照)。このプログラムは予測されたエキソン群を連結し、メチオニンから終止コドンに及ぶ、構築されたcDNA配列を形成する。Genscanの出力は、ポリヌクレオチドおよびポリペプチド配列のFASTAデータベースである。Genscanが一度に分析する配列の最大範囲は、30 kbに設定した。これらのGenscan予測cDNA配列の内、どの配列が細胞成長、分化および細胞死関連タンパク質をコードするかを判定するために、コードされるポリペプチドを、PFAMモデル群に対し細胞成長、分化および細胞死関連タンパク質について問合せて分析した。潜在的な細胞成長、分化および細胞死関連タンパク質はまた、既に細胞成長、分化および細胞死関連タンパク質としてアノテーションが付けられたIncyte cDNA配列に対する相同性を基に同定された。これら選択したGenscan予測配列を、次にBLAST分析により、公共データベースgenpeptおよびgbpriと比較した。必要であれば、genpeptからのトップBLASTヒットと比較することによりGenscan予測配列を編集し、余分なまたは省略されたエキソンなど、Genscanが予測した配列におけるエラーを補正した。BLAST分析はまた、Genscan予測配列の、いかなるIncyte cDNAまたは公共cDNAカバレッジ (coverage) の発見にも用いられ、したがって転写の証拠を提供した。Incyte cDNAカバレッジが利用できた場合には、この情報を用いてGenscan予測配列を補正または確認した。完全長ポリヌクレオチド配列は、実施例3に記載した構築プロセスを用いて、Incyte cDNA配列および/または公共cDNA配列でGenscan予測コード配列を構築して得た。あるいは、完全長ポリヌクレオチド配列は、編集後または非編集のGensca

30

40

50

n予測コード配列に、完全に由来する。

【0383】

5 ゲノム配列データのcDNA配列データとの統合

スティッチ配列 (Stitched Sequence)

部分cDNA配列群を伸長させるため、実施例 4に記載のGenscan遺伝子同定プログラムが予測したエキソン群を用いた。実施例 3に記載したように構築した部分cDNA群を、ゲノムDNAにマッピングし、また、関連するcDNA群と、1つ以上のゲノム配列からGenscan予測されたエキソン群とを有するクラスター群に分解した。cDNAとゲノムとの情報を統合すべく、グラフ理論および動的プログラミングに基づく或るアルゴリズムを用いて各クラスターを分析し、潜在的スプライス変異体群を生成した。配列群を続いて確認、編集または伸長し、完全長配列を創出した。区間の全長が、2つ以上の配列に在るような配列区間群をクラスター内で同定し、そのように同定された区間群は、推移性により、等しいと考えた。例えば、或る区間が、1つのcDNAと2つのゲノム配列とに在る場合、3つの区間は全て等しいと考えた。このプロセスにより、無関係だが連続したゲノム配列群を、cDNA配列で結び合わせて架橋し得る。このようにして同定した区間を、それらの親配列 (parent sequences) に沿って現われる順にスティッチアルゴリズムで「縫い合わせ」、可能な限り最長の配列と配列変異体群とを生成した。1種類の親配列に沿って続く区間間の連鎖 (cDNA-cDNAまたはゲノム配列-ゲノム配列) は、親の種類を変える連鎖 (cDNA-ゲノム配列) より優先した。結果として得たスティッチ配列群を翻訳し、BLAST分析で公共データベースgenpeptおよびgbpriと比較した。Genscanが予測した不正確なエキソン群は、genpeptからのトップBLASTヒットとの比較により補正した。必要な場合、追加cDNA配列群を用いるかゲノムDNAの検査により、配列群を更に伸長させた。

10

20

【0384】

ストレッチ配列 (Stretched Sequence)

部分DNA配列群を、BLAST分析に基づく1アルゴリズムにより完全長まで伸長した。先ず、BLASTプログラムを用いて、GenBankの霊長類、げっ歯類、哺乳類、脊椎動物および真核生物のデータベースなどの公共データベースに対し、実施例 3に記載されたように構築された部分cDNAを問い合わせた。次に、最も近いGenBankタンパク質相同体を、BLAST分析により、Incyte cDNA配列または実施例 4に記載のGenScanエキソン予測配列のいずれかと比較した。結果として得られる高スコアリングセグメント対 (HSP) を用いてキメラタンパク質を産出し、翻訳した配列をGenBankタンパク質相同体上にマッピングした。元のGenBankタンパク質相同体に対し、キメラタンパク質内では挿入または欠失が起こり得る。GenBankタンパク質相同体、キメラタンパク質またはその両方をプローブとして用い、公共のヒトゲノムデータベースから相同ゲノム配列を検索した。このようにして、部分的なDNA配列を、相同ゲノム配列の付加によりストレッチすなわち伸長した。結果として得られるストレッチ配列を検査し、完全遺伝子を含んでいるか否かを判定した。

30

【0385】

6 CGDDをコードするポリヌクレオチドの染色体マッピング

SEQ ID NO:22-42を構築するために用いた配列群を、BLAST他のSmith-Watermanアルゴリズムの実装群を用いて、Incyte LIFESEQデータベースおよび公共ドメインデータベース群の配列と比較した。SEQ ID NO:22-42と一致するこれらのデータベースの配列を、Phrapなどの構築アルゴリズム (表7) を使用して、連続しオーバーラップする配列のクラスター群にアセンブリした。スタンフォード・ヒトゲノムセンター (SHGC)、ホワイトヘッド・ゲノム研究所 (WIGR)、Genethonなどの公的な情報源から入手可能な放射線ハイブリッドおよび遺伝地図データを用いて、いずれかのクラスター化された配列が既にマッピングされているかを判定した。マッピングされた配列が或るクラスターに含まれている場合、そのクラスターの全配列が、個々の配列番号と共に、地図上の位置に割り当てられた。

40

【0386】

地図上の位置は、ヒト染色体の範囲または区間として表される。或る区間の地図上の位置 (センチモルガン単位) は、染色体のpアームの末端に対して測定する (センチモルガン (c

50

M) は、染色体マーカー間の組換え頻度に基づく計測単位である。平均して、1 cMは、ヒトDNAの1メガベース (Mb) にほぼ等しい。尤も、この値は、組換えのホットスポットおよびコールドスポットに起因して広範囲に変化する)。cM距離は、各クラスタ内に配列が含まれる放射線ハイブリッドマーカー類に対して境界を提供するGenethonによってマッピングされた遺伝マーカー群に基づく。NCBI「GeneMap'99」 (<http://www.ncbi.nlm.nih.gov/genemap/>) など一般個人が入手可能なヒトゲノム地図などの資源を用いて、既に同定された疾患遺伝子類が、上記の区間内若しくは近傍にマップされているかを判定できる。

【0387】

この方法で、SEQ ID NO:26は、3番染色体の63.30~77.40センチモルガンの区間内へマッピングされた。

【0388】

7 ポリヌクレオチド発現の分析

ノーザン分析は、遺伝子の転写物の存在を検出するために用いられる実験技術であり、特定の細胞種または組織からのRNAが結合している膜への、標識されたヌクレオチド配列のハイブリダイゼーションに関与する (例えば前出のSambrook, 7章、同Ausubel (1995) 4章および16章を参照)。

【0389】

類似の、BLASTを応用したコンピュータ技術を用いて、GenBankやLIFESEQ (Incyte Genomics) などのcDNAデータベースにおいて、同一または関連分子を検索した。ノーザン分析は、いくつかの膜系ハイブリダイゼーションよりも非常に速い。更に、任意の特定の一致を厳密なあるいは相同的なものとして分類するか否かを決定するため、コンピュータ検索の感度を修正することができる。検索の基準は積スコアであり、次式で定義される。

【0390】

【数1】

(BLASTスコア×配列一致率)

5 × (長さ(配列1), 長さ(配列2))の最小値

【0391】

積スコアは、2つの配列間の類似度と、配列が一致する長さとの両方を考慮している。積スコアは0~100のノーマライズされた値であり、次のようにして求める。BLASTスコアにヌクレオチドの配列一致率を乗じ、その積を2つの配列の短い方の長さの5倍で除する。BLASTスコアを計算するには、或る高スコアリングセグメント対 (HSP) 内の一致する各塩基に+5のスコアを割り当て、各不一致塩基に-4を割り当てる。2つの配列は、2以上のHSPを共有し得る (ギャップにより隔離される)。2以上のHSPがある場合には、最高BLASTスコアのセグメント対を用いて積スコアを計算する。積スコアは、断片的オーバーラップとBLASTアラインメントの質とのバランスを表す。例えば積スコア100は、比較した2つの配列の短い方の長さ全体にわたって100%一致する場合のみ得られる。積スコア70は、一端が100%一致し、70%オーバーラップしているか、他端が88%一致し、100%オーバーラップしているかのいずれかの場合に得られる。積スコア50は、一端が100%一致し、50%オーバーラップしているか、79%一致し、100%オーバーラップしているかのいずれかの場合に得られる。

【0392】

CGDDをコードするポリヌクレオチド配列を別法では、由来する組織に対して分析する。例えば幾つかの完全長配列は、少なくとも一部は、オーバーラップするIncyte cDNA配列群を用いて構築される (実施例3を参照)。各cDNA配列は、ヒト組織から作製されたcDNAライブラリに由来する。各ヒト組織は、以下の臓器/組織カテゴリーの1つへ分類される。すなわち心血管系、結合組織、消化器系、胎芽構造、内分泌系、外分泌腺、女性生殖器、男性生殖器、生殖細胞、血液および免疫系、肝、筋骨格系、神経系、脾臓、呼吸器系、感覚器、皮膚、顎口腔系、非分類性/混合性または尿路である。各カテゴリーのライブラリ数

10

20

30

40

50

を数えて、全カテゴリーの総ライブラリ数で除する。同様に、各ヒト組織は、以下の疾患／条件カテゴリー、すなわち癌、細胞株、発達、炎症、神経性、外傷、心血管、プール、その他の1つに分類される。各カテゴリーのライブラリ数を数えて、全カテゴリーの総ライブラリ数で除する。CGDDをコードするcDNAの疾患特異的な発現を、得られるパーセンテージが反映する。cDNA配列群およびcDNAライブラリ／組織の情報は、LIFESEQ GOLDデータベース (Incyte Genomics, Palo Alto CA) から得ることができる。

【0393】

8 CGDDをコードするポリヌクレオチドの伸長

完全長のポリヌクレオチド配列はまた、完全長分子の適切な断片から設計したオリゴヌクレオチドプライマー群を用いて該断片を伸長させて産生した。或るプライマーは既知の断片の5'伸長を開始するべく合成し、別のプライマーは既知の断片の3'伸長を開始するべく合成した。開始プライマー群の設計にはOLIGO 4.06ソフトウェア (National Biosciences) 或いは別の適切なプログラムを用い、長さが約22～30ヌクレオチド、GC含有率が約50%以上となり、約68～約72℃の温度で標的配列にアニーリングするようにした。ヘアピン構造およびプライマー-プライマー二量体を生ずるようなヌクレオチド群の伸長は、全て回避した。

【0394】

選択したヒトcDNAライブラリ群を用い、配列を伸長した。2段階以上の伸長が必要または望ましい場合には、付加的プライマーあるいはプライマーのネステッドセットを設計した。

【0395】

高忠実度の増幅が、当業者によく知られている方法を利用したPCR法によって得られた。PCRは、PTC-200 サーマルサイクラー (MJ Research, Inc.) を用いて96穴プレート内で行った。反応混液は、鋳型DNAを有し、また、200 nmolの各プライマーを有する。また、 Mg^{2+} と $(NH_4)_2SO_4$ と2-メルカプトエタノールとを含有する反応バッファーと、Taq DNAポリメラーゼ (Amersham Pharmacia Biotech) と、ELONGASE酵素 (Life Technologies) と、Pfu DNAポリメラーゼ (Stratagene) とを含有する。プライマー対であるPCI AとPCI Bとに対し、以下のパラメータで増幅した。ステップ1: 94℃で3分間。ステップ2: 94℃で15秒間。ステップ3: 57℃で1分間。ステップ4: 68℃で2分間。ステップ5: ステップ2、3および4を20回繰り返す。ステップ6: 68℃で5分間。ステップ7: 4℃で保存。

【0396】

各ウェルのDNA濃度は、1×TEおよび0.5μlの希釈していないPCR産物に溶解した100μlのPICOGREEN定量試薬(0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR)を不透明な蛍光光度計プレート(Corning Costar, Acton MA)の各ウェルに分配し、DNAが試薬と結合できるようにして判定した。サンプルの蛍光を計測してDNAの濃度を定量すべく、プレートをFluoroskan II (Labsystems Oy, Helsinki, Finland) でスキャンした。反応混合物のアリコット5～10μlを1%アガロースゲル上で電気泳動法によって解析し、どの反応が配列の伸長に成功したかを判定した。

【0397】

伸長したヌクレオチドは、脱塩および濃縮して384穴プレートに移し、CviJIコレラウイルスエンドヌクレアーゼ (Molecular Biology Research, Madison WI) を用いて消化し、音波処理またはせん断し、pUC 18ベクター (Amersham Pharmacia Biotech) への再連結を行った。ショットガン・シークエンシングのために、消化したヌクレオチドを低濃度(0.6～0.8%)のアガロースゲル上で分離し、断片を切除し、寒天をAgar ACE (Promega) で消化した。伸長させたクローンをT4リガーゼ (New England Biolabs, Beverly MA) を用いてpUC 18ベクター (Amersham Pharmacia Biotech) に再連結し、Pfu DNAポリメラーゼ (Stratagene) で処理して制限部位のオーバーハングを満たし、適格な大腸菌細胞に形質移入した。形質移入した細胞群の選択を、抗生物質を含む培地で行い、それぞれのコロニーを採取し、LB/2Xカルベニシリン培養液中の384ウェルプレート群に37℃で一晩培養した。

【0398】

細胞を溶解し、Taq DNAポリメラーゼ (Amersham Pharmacia Biotech) およびPfu DNAポリメラーゼ (Stratagene) を用いて以下のパラメータでDNAをPCR増幅した。ステップ1: 94℃で3分間。ステップ2: 94℃で15秒間。ステップ3: 60℃で1分間。ステップ4: 72℃で2分間。ステップ5: ステップ2、3および4を29回繰り返す。ステップ6: 72℃で5分間。ステップ7: 4℃で保存。DNAの定量化には、上記したようにPICOGREEN試薬 (Molecular Probes) を用いた。DNAの回収率が低いサンプルは、上記と同一の条件を用いて再増幅した。サンプルは20%ジメチルスルホキシド (1:2, v/v) で希釈し、DYENAMICエネルギー移動シークエンシングプライマー、およびDYENAMIC DIRECT kit (Amersham Pharmacia Biotech) またはABI PRISM BIGDYEターミネーターサイクル シークエンシングレディ反応キット (Terminator cycle sequencing ready reaction kit) (Applied Biosystems) を用いてシークエンシングした。 10

【0399】

同様に、上記手順を用いて完全長ポリヌクレオチド配列を検証した。あるいは、完全長ポリヌクレオチドを用い、上記手順で、そのような伸長のために設計したオリゴヌクレオチド類と、或る適切なゲノムライブラリとを用いて5'調節配列を得た。

【0400】

9 CGDDをコードするポリヌクレオチドにおける1塩基多型の同定

1塩基多型 (SNP) として知られる一般的なDNA配列変異体が、SEQ ID NO:22-42において、LIFESEQデータベース (Incyte Genomics) を用いて同定された。同じ遺伝子からの配列群は共にクラスター化され、実施例3に述べたように構築されて、その遺伝子における全ての配列変異体の同定を可能にした。一連のフィルタ群を有する或るアルゴリズムを用いて、SNPを他の配列変異体から区別した。前段フィルタ群 (preliminary filters) が過半の塩基呼び出し (basecall) エラーの除去を、最少Phred質スコア15を要求することによって行い、また、配列アラインメントエラーと、ベクター配列、キメラ、およびスプライス変異体の不適切なトリミングの結果であるエラーとをも除去した。先進の染色体分析の自動化された手順によって、推定上のSNPの近傍にある、オリジナルのクロマトグラムファイル群を分析した。クローンエラーフィルタ群は、統計的に発生させたアルゴリズムを用いて、実験室でのプロセス中に生じたエラーを同定した。エラーとしては、逆転写酵素、ポリメラーゼ、または常染色体突然変異によって起きたエラーがある。クラスター化エラーフィルタ群は、統計的に発生させたアルゴリズムを用いて、近縁の相同体群または偽遺伝子群のクラスター化の結果であるエラーや、非ヒト配列による汚染が原因のエラーを同定した。最終セットのフィルタ群は、免疫グロブリンまたはT細胞受容体において見られる重複 (duplicates) とSNPとを除去した。 20 30 40

【0401】

幾つかのSNPは、更なる特徴付けの為に選択された。選択は高処理MASSARRAYシステム (Sequenom, Inc.) を用いた質量分析によって行い、4群の異なるヒト集団におけるSNP部位での対立遺伝子頻度を分析した。白人集団は92個体 (46名の男子、46名の女子) からなり、83名はユタ出身、4名はフランス人、3名はベネズエラ人、2名はアーミッシュの人々である。アフリカ集団は194個体 (97名の男子、97名の女子) からなり、全員がアフリカ系アメリカ人である。ヒスパニック集団は324個体 (162名の男子、162名の女子) からなり、全員がメキシコのヒスパニックである。アジア集団は126個体 (64名の男子、62名の女子) からなり、報告された親の内訳は43%が中国人、31%が日本人、13%がコリアン、5%がベトナム人、8%が他のアジア人である。対立遺伝子頻度の分析は先ず白人集団においてなされ、幾つかの場合、SNPの内、対立遺伝子変異をこの集団において示さなかったSNPは、更に他の3集団においてテストされることは無かった。 40

【0402】

10 個々のハイブリダイゼーションプローブの標識化および使用

SEQ ID NO:22-42由来のハイブリダイゼーションプローブを用いて、cDNA、mRNA、またはゲノムDNAをスクリーニングする。約20塩基対からなるオリゴヌクレオチドの標識につい 50

て特に記載するが、より大きなヌクレオチド断片に対しても本質的に同一の手順が用いられる。オリゴヌクレオチドは、OLIGO 4.06ソフトウェア (National Biosciences) 等の最新ソフトウェアを用いて設計し、各オリゴマー 50 pmolと、 $[\gamma\text{-}^{32}\text{P}]$ アデノシン 3 リン酸 (Amersham Pharmacia Biotech) 250 μCi と、T4ポリヌクレオチドキナーゼ (DuPont NEN, Boston MA) とを混合することにより標識する。標識したオリゴヌクレオチドは、SEPHADEX G-25超細繊維分子サイズ排除デキストラン ビーズカラム (Amersham Pharmacia Biotech) を用いて実質的に精製する。Ase I、Bgl II、Eco RI、Pst I、Xba IまたはPvu II (DuPont NEN) のいずれか1つのエンドヌクレアーゼで消化されたヒトゲノムDNAの、典型的な膜ベースのハイブリダイゼーション解析において、毎分 10^7 カウントの標識されたプローブを含むアリコットを用いる。

10

【0403】

各消化物から得たDNAは、0.7%アガロースゲル上で分画してナイロン膜 (Nytran Plus, Schleicher & Schuell, Durham NH) に移す。ハイブリダイゼーションは、40°Cで16時間行う。非特異的シグナル群を除去するため、最大で例えば0.1×クエン酸ナトリウム食塩水および0.5%ドデシル硫酸ナトリウムの条件下で、ブロット群を室温で順次洗浄する。オートラジオグラフィーまたはそれに代わるイメージング手段を用いてハイブリダイゼーションパターンを視覚化し、比較する。

【0404】

11 マイクロアレイ

或るマイクロアレイ上でのアレイエレメント群の接続または合成は、フォトリソグラフィ、ピエゾ式印刷 (インクジェット印刷、前出のBaldeschweilerなどを参照)、機械的マイクロスポッティング技術およびこれらから派生した技術を用いて達成し得る。上記各技術において基板は、均一な、無孔の表面を持つ固体とすべきである (Skena (1999) 前出)。推奨する基板には、シリコン、シリカ、スライドガラス、ガラスチップおよびシリコンウェハがある。あるいは、ドットブロット法またはスロットブロット法に類似した手順を利用し、熱的、紫外線的、化学的または機械的結合手順を用いて基板の表面にエレメントを配置および結合させてもよい。通常のアレイは、利用可能な、当業者に公知の方法と機械とを用いて作製でき、任意の適正数のエレメントを有し得る (例えばSkena, M. 他 (1995) Science 270:467-470; Shalon, D. 他 (1996) Genome Res. 6:639-645; Marshall, A. および J. Hodgson (1998) Nat. Biotechnol. 16:27-31を参照)。

20

30

【0405】

完全長cDNA、発現配列タグ (EST)、またはそれらの断片またはオリゴマーが、マイクロアレイのエレメントと成り得る。ハイブリダイゼーションに好適な断片またはオリゴマーを、LASERGENEソフトウェア (DNASTAR) など本技術分野で公知のソフトウェアを用いて選択することが可能である。アレイエレメント群を、生体サンプル中のポリヌクレオチド群とハイブリダイズする。生体サンプル中のポリヌクレオチドは、検出を容易にするために蛍光標識などの分子タグに抱合させる。ハイブリダイゼーション後、生体サンプルからのハイブリダイズされていないヌクレオチドを除去し、蛍光スキャナを用いて各アレイエレメントでのハイブリダイゼーションを検出する。あるいは、レーザー脱離および質量スペクトロメトリを用いてもハイブリダイゼーションを検出し得る。マイクロアレイ上の或るエレメントにハイブリダイズする各ポリヌクレオチドの、相補性の度合と相対存在度とを算定し得る。一実施態様におけるマイクロアレイの調製および使用について、以下に詳述する。

40

【0406】

組織または細胞サンプルの調製

全RNAを組織サンプル群から単離するためグアニジニウム チオシアネート法を用い、ポリ (A)⁺ RNAを精製するためオリゴ (dT) セルロース法を用いる。各ポリ (A)⁺ RNAサンプルを、MLV逆転写酵素、0.05pg/ μl のオリゴ (dT) プライマー (21mer)、1×第一鎖バッファー、0.03unit/ μl のRNアーゼ阻害因子、500 μM のdATP、500 μM のdGTP、500 μM のdTTP、40 μM のdCTP、40 μM のdCTP-Cy3 (BDS) またはdCTP-Cy5 (Amersham Pharmacia Biotech

50

）を用いて逆転写する。逆転写反応は、GEMBRIGHTキット（Incyte）を用い、200 ngのポリ（A）⁺RNAを含有する体積25 mlで行う。特異的対照ポリ（A）⁺RNA群は、非コード酵母ゲノムDNAからin vitro転写により合成する。37℃で2時間インキュベートした後、各反応サンプル（1つはCy3、もう1つはCy5標識）は、2.5 mlの0.5 M水酸化ナトリウムで処理し、85℃で20分間インキュベートし、反応を停止させてRNAを分解させる。サンプル群の精製には、2つの連続するCHROMA SPIN 30ゲル濾過スピンカラム（CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA）を用いる。混合後、2つの反応サンプルのエタノール沈殿を、1 mlのグリコーゲン（1 mg/ml）、60 mlの酢酸ナトリウム、および300 mlの100%エタノールで行う。サンプルは次に、SpeedVAC（Savant Instruments Inc., Holbrook NY）を用いて完全に乾燥させ、14 μlの5×SSC/0.2% SDS中で再懸濁する。 10

【0407】

マイクロアレイの調製

本発明の配列を用いて、アレイエレメントを作製する。各アレイエレメントは、クローン化したcDNAインサート群と、ベクター群とを含有する細菌細胞群から増幅する。PCR増幅は、cDNAインサートの側面に位置するベクター配列群に相補的なプライマー類を用いる。30サイクルのPCRによって、1～2 ngの初期量から5 μgを超える最終量まで、アレイエレメントを増幅する。増幅したアレイエレメントは、SEPHACRYL-400（Amersham Pharmacia Biotech）を用いて精製する。

【0408】

精製したアレイエレメントは、ポリマーコートされたスライドガラス上に固定する。顕微鏡スライドガラス（Corning）は、0.1%のSDSおよびアセトン中で超音波洗浄し、処理の間および処理後に十分に蒸留水で洗浄する。スライドガラスは、4%フッ化水素酸（VWR Scientific Products Corporation (VWR), West Chester PA）中でエッチングし、十分に蒸留水中で洗浄し、95%エタノール中で0.05%アミノプロピルシラン（Sigma）を用いてコーティングする。コーティングしたスライドは、110℃のオーブンで硬化させる。

【0409】

米国特許第5,807,522号に記載されている手順を用いて、コーティングしたガラス基板にアレイエレメント群を付加する。この特許は、引用を以って本明細書の一部とする。平均濃度100 ng/μlのアレイエレメントDNA 1 μlを、高速ロボット装置（robotic apparatus）により、開放型キャピラリープリンティングエレメント（open capillary printing element）に充填する。装置はここで、スライド毎に約5 nlのアレイエレメントサンプルを置く。 30

【0410】

マイクロアレイをUV架橋するため、STRATALINKER UV架橋剤（Stratagene）を用いる。マイクロアレイの洗浄を、室温において、0.2% SDSで1回、蒸留水で3回行う。非特異結合部位をブロックするため、マイクロアレイのインキュベーションをリン酸緩衝生理食塩水（PBS）（Tropix, Inc., Bedford MA）中の0.2%カゼイン中において60℃で30分間行った後、前に行ったように0.2% SDSおよび蒸留水で洗浄する。 40

【0411】

ハイブリダイゼーション

ハイブリダイゼーション反応に用いる9 μlのサンプル混合体には、Cy3またはCy5で標識したcDNA合成産物群の各0.2 μgを、5×SSC, 0.2% SDSハイブリダイゼーション緩衝液中に含む。サンプル混合体は、65℃まで5分間加熱し、マイクロアレイ表面上へ等分して1.8 cm²のカバーガラスで覆う。アレイを、顕微鏡用スライドよりわずかに大きい空洞を有する防水チェンバーに移す。チェンバー内を湿度100に保つため、140 μlの5×SSCをチェンバーの1コーナーに加える。アレイを入れたチェンバーは、60℃で約6.5時間インキュベートする。アレイは、第1洗浄緩衝液中（1×SSC, 0.1% SDS）において45℃で10分間洗浄し、第2洗浄緩衝液中（0.1×SSC）において45℃で10分間ず 50

つ3回洗浄し、乾燥させる。

【0412】

検出

レポーター標識したハイブリダイゼーション複合体を検出するには、Cy3の励起のために488nm、Cy5の励起のために632nmでスペクトル線を発生し得るInnova70混合ガス10Wレーザー (Coherent, Inc., Santa Clara CA) を備えた顕微鏡を用いる。励起レーザー光の焦点をアレイ上に置くため、20×顕微鏡対物レンズ (Nikon, Inc., Melville NY) を用いる。アレイを含むスライドを、顕微鏡の、コンピュータ制御のX-Yステージに置き、対物レンズを通してラスタースキャンする。本実施例で用いる1.8cm×1.8cmのアレイは、解像度20μmでスキャンする。

10

【0413】

異なる2回のスキャンで、混合ガスマルチラインレーザーは2つのフルオロフォアを連続的に励起する。発光された光は、波長に基づき分離され、2つのフルオロフォアに対応する2つの光電子増倍管検出器 (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) に送られる。好適なフィルタ群をアレイと光電子増倍管との間に設置して、シグナルをフィルタする。用いるフルオロフォアの最大発光の波長は、Cy3では565nm、Cy5では650nmである。装置は両方のフルオロフォアからのスペクトルを同時に記録し得るが、レーザー源において好適なフィルタを用いて、フルオロフォア1つにつき1度スキャンし、各アレイを通常2度スキャンする。

20

【0414】

通常、各スキャンの感度を校正するため、cDNA対照種を或る既知濃度でサンプル混合体に添加し、対照種が発生するシグナル強度を用いる。アレイ上の或る特定の位置には或る相補的DNA配列が含まれ、その位置におけるシグナルの強度をハイブリダイゼーション種の重量比1:100,000で相関させる。異なる源泉 (例えば代表的な試験細胞および対照細胞など) からの2つのサンプルを、各々異なるフルオロフォアで標識し、異なる発現をする遺伝子群を同定するために単一のアレイにハイブリダイズする場合には、校正するcDNAのサンプルを2種のフルオロフォアで標識し、ハイブリダイゼーション混合液に各々等量を加えることによって校正を行う。

【0415】

光電子増倍管の出力をデジタル化するため、IBMコンパチブルPCコンピュータにインストールした12ビットRTI-835Hアナログデジタル (A/D) 変換ボード (Analog Devices, Inc., Norwood MA) を用いる。デジタル化したデータは、青色 (低シグナル) から赤色 (高シグナル) までの擬似カラースケールへのリニア20色変換を用いて、シグナル強度がマッピングされたイメージとして表示される。データは、定量的にも分析される。2つの異なるフルオロフォアを同時に励起および測定する場合には、各フルオロフォアの発光スペクトルを用いて、データは先ずフルオロフォア間の光学的クロストーク (発光スペクトルの重なりに起因する) を補正される。

30

【0416】

グリッドが蛍光シグナルイメージ上に重ねられ、それによって各スポットからのシグナルはグリッドの各エレメントに集められる。各エレメント内の蛍光シグナルは統合され、シグナルの平均強度に応じた数値が得られる。シグナル分析に用いるソフトウェアは、GEMT00LS遺伝子発現分析プログラム (Incyte) である。

40

【0417】

発現

正常乳房細胞株を、次のようにして得る。初代乳腺細胞の単離を、線維嚢胞性乳房疾患 (fibrocystic breast disease) のドナーから行う。または、初代乳腺上皮細胞の単離を、正常ドナーから行う。乳癌細胞の由来は、in vitroの、腫瘍から遊出した細胞である。正常および、種々のステージの腫瘍化乳房細胞株の購入を、American Type Culture Collection (ATCC), (Manassas, VA) から行った。

【0418】

50

例えばSEQ ID NO:34 は、癌細胞株や腫瘍組織では、非癌細胞株または組織に対して差次的発現を示し、これはマイクロアレイ分析で判定した。CGDD-13の発現は、少なくとも3倍増が、或る乳腺腫瘍細胞株で見られた。この株の取得ドナーは、腫瘍進行の早期にあった。

【0419】

別の例でSEQ ID NO:37は差次的発現を炎症応答で示し、これはマイクロアレイ分析で判定した。SEQ ID NO:37 の発現は、少なくとも2倍増が、LPS処理したPBMCで見られた。これは無処理PBMCと比較したものである。したがってSEQ ID NO:37は、炎症応答の診断アッセイに有用である。

【0420】

またSEQ ID NO:37は、マイクロアレイ分析の判定で、様々な乳癌株と非悪性乳腺上皮細胞とで差次的発現を示した。SEQ ID NO:37 の発現は、非悪性乳腺上皮細胞に比べて乳癌株で少なくとも2倍減少していた。したがってSEQ ID NO:37 は、乳癌検知の診断アッセイに有用である。

【0421】

別の例でSEQ ID NO:41 はアルツハイマー病患者の脳帯状回と、同じドナーのマッチした微視的正常組織とで差次的発現を示し、これはマイクロアレイ分析で判定した。CGDD-19の発現の増加が、アルツハイマー病の帯状組織に見られた。したがってSEQ ID NO:41は、神経疾患、特にアルツハイマー病の診断アッセイに有用である。

【0422】

1.2 相補的ポリヌクレオチド

CGDDをコードする配列或いはその任意の一部に対して相補的な配列は、天然CGDDの発現を検出、低減または阻害するために用いられる。約15~30塩基対を含むオリゴヌクレオチドの使用について記すが、これより小さなあるいは大きな配列の断片の場合でも、本質的に同じ手順を用いる。CGDDのコーディング配列とOligo4.06ソフトウェア (National Biosciences) とを用いて、適切なオリゴヌクレオチドを設計する。転写を阻害するためには、最も独特な5'配列から相補的オリゴヌクレオチドを設計し、これを用いて、プロモーターがコーディング配列に結合するのを防止する。CGDDをコードする転写物にリボソームが結合するのを阻害するよう相補的なオリゴヌクレオチドを設計すると、翻訳を阻害する。

【0423】

1.3 CGDDの発現

CGDDの発現および精製は、細菌若しくはウイルスを基にした発現系を用いて行うことができる。CGDDを細菌内で発現させるには、抗生物質耐性遺伝子と、cDNAの転写レベルを高める誘導性プロモーターとを持つ、好適なベクターに、cDNAをサブクローニングする。このようなプロモーターとしては、lacオペレーター調節エレメントと併用するT5またはT7バクテリオファージプロモーター、およびtrp-lac (tac) ハイブリッドプロモーターが含まれるが、これらに限定するものではない。組換えベクターを、BL21 (DE3) などの好適な細菌宿主に形質転換する。CGDDを抗生物質耐性細菌に発現させるには、イソプロピルβ-Dチオガラクトピラノシド (IPTG) で誘発する。CGDDの真核細胞での発現は、昆虫細胞株または哺乳動物細胞株に、一般にバキュロウイルスとして知られる *Autographica californica* 核多角体病ウイルス (AcMNPV) の組換え型を感染させて行う。CGDDをコードするcDNAとバキュロウイルスの非必須ポリヘドリン遺伝子を置換するには、相同組換えを行うか、或いは、転移プラスミドの媒介を伴う、細菌の媒介による遺伝子転移を行う。ウイルスの感染力は維持され、強力なポリヘドリンプロモーターによって高レベルのcDNA転写が行われる。組換えバキュロウイルスは、多くの場合は夜蛾の1種 *Spodoptera frugiperda* (Sf9) 昆虫細胞への感染に用いるが、ヒト肝細胞への感染に用いることもある。後者の感染の場合は、バキュロウイルスへの更なる遺伝的修飾が必要になる (Engelhard, E.K.他 (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V.他 (1996) *Hum. Gene Ther.* 7:1937-1945)。

【0424】

CGDDは殆どの発現系で、例えばグルタチオンSトランスフェラーゼ(GST)、またはFLAGや6-Hisなどのペプチドエピトープ標識で合成された融合タンパク質となるため、未精製の細胞溶解物からの組換え融合タンパク質の親和性ベースの精製を、迅速に1回で行うことができる。GSTは日本住血吸虫からの26kDaの酵素であり、タンパク質の活性および抗原性を維持した状態で、固定化したグルタチオン上での融合タンパク質の精製を可能とする(Amersham Pharmacia Biotech)。CGDDから、精製後にGST部分を、特異的に操作した諸部位においてタンパク分解的に切断できる。FLAGは8アミノ酸のペプチドであり、市販されているモノクローナルおよびポリクローナル抗FLAG抗体(Eastman Kodak)を用いた免疫親和性精製を可能にする。6ヒスチジン残基が連続して伸長した6-Hisは、金属キレート樹脂上での精製を可能にする(QIAGEN)。タンパク質の発現および精製の方法は、前出のAusubel(1995)10章、16章に記載されている。CGDDをこれらの方法で精製し直接用いて、以下の実施例17、18及び19の、適用可能なアッセイを行い得る。

【0425】

1.4 機能的アッセイ

CGDD機能は、CGDDをコードする配列群の、哺乳動物細胞培養系において生理的に高められたレベルでの発現によって算定する。cDNAを、cDNAを高いレベルで発現させる強いプロモーターを持つ哺乳類発現ベクターにサブクローニングする。選択されるベクターとしては、PCMV SPORT (Life Technologies) およびPCR 3.1 (Invitrogen, Carlsbad CA) があり、どちらもサイトメガロウイルスプロモーターを持つ。リポソーム製剤あるいは電気穿孔法を用いて、5~10 μ gの組換えベクターをヒト細胞株、例えば内皮由来または造血由来の細胞株に、一過的に形質移入する。更に、標識タンパク質をコードする配列を含む1~2 μ gのプラスミドを同時に形質移入する。標識タンパク質の発現により、形質移入細胞と非形質移入細胞を区別する手段が与えられる。また、標識タンパク質の発現によって、組換えベクターからのcDNA発現を正確に予想できる。標識タンパク質は、例えば緑色蛍光タンパク質(GFP; Clontech)、CD64またはCD64-GFP融合タンパク質から選択できる。自動化された、レーザー光学に基づく技術であるフローサイトメトリー(FCM)を用いて、GFPまたはCD64-GFPを発現する形質移入された細胞を同定し、それらの細胞のアポトーシス状態や他の細胞特性を評価する。FCMは、細胞死に先行するかあるいは同時に発生する現象を診断する蛍光分子の取込を検出して計量する。このような現象として挙げられるのは、ヨウ化プロピジウムによるDNA染色によって計測される核DNA含量の変化、前方散乱光と90°側方散乱光によって計測される細胞サイズと顆粒性の変化、プロモデオキシウリジンの取込量の低下によって計測されるDNA合成の下方調節、特異抗体との反応性によって計測される細胞表面および細胞内におけるタンパク質の発現の変容、および、フルオレセイン抱合したアネキシンVタンパク質の細胞表面への結合によって計測される原形質膜組成の変容がある。フローサイトメトリー法については、Ormerod, M.G. (1994) *Flow Cytometry*, Oxford, New York NYに記述がある。

【0426】

CGDDの遺伝子発現への影響は、CGDDをコードする配列と、CD64またはCD64-GFPのどちらかが形質移入された、高度に精製された細胞集団を用いて評価することができる。CD64またはCD64-GFPは、形質移入された細胞表面で発現し、ヒト免疫グロブリンG(IgG)の保存された複数の領域に結合する。形質移入された細胞と形質移入されない細胞とは、ヒトIgGかCD64に対する抗体のどちらかで被覆された磁気ビーズを用いて効率よく分離できる(DYNAL, Lake Success NY)。mRNAは、これら細胞から、当業者に周知の方法で精製できる。CGDDと目的の他の遺伝子群とをコードするmRNAの発現は、ノーザン分析やマイクロアレイ技術で分析できる。

【0427】

1.5 CGDDに特異的な抗体の産生

CGDDの実質的精製を、ポリアクリルアミドゲル電気泳動法(PAGE; 例えばHarrington, M. G. (1990) *Methods Enzymol.* 182:488-495を参照)または他の精製技術で行い、これを用いて標準的なプロトコルで動物(例えばウサギ、マウスなど)を免疫化して抗体を作り出

す。

【0428】

CGDDアミノ酸配列を別法ではLASERGENEソフトウェア (DNASTAR) を用いて解析して免疫原性の高い領域を決定し、対応するオリゴペプチドを合成してこれを用いて当業者に周知の方法で抗体を生産する。C末端付近の、或いは隣接する親水性領域内のエピトープなどの適切なエピトープの選択については、当分野で周知である (例えば、前出のAusubel, 1995, 11章を参照)。

【0429】

通常は、長さ約15残基のオリゴペプチドを、FMOCケミストリを用いるABI 431A ペプチドシンセサイザ (Applied Biosystems) を用いて合成し、N-マレイミドベンゾイル-N-ヒドロキシスクシンイミドエステル (MBS) を用いた反応によってKLH (Sigma-Aldrich, St. Louis MO) に結合させて、免疫原性を高める (例えば前出のAusubel, 1995を参照)。完全フロイントアジュバントにおいて、オリゴペプチド-KLH複合体を用いてウサギを免疫化する。得られた抗血清の抗ペプチド活性および抗CGDD活性を検査するには例えば、ペプチドまたはCGDDを基板に結合し、1% BSAを用いてブロッキング処理し、ウサギ抗血清と反応させて洗浄し、さらに、放射性ヨウ素標識したヤギ抗ウサギIgGと反応させる。

【0430】

1.6 特異抗体を用いる天然CGDDの精製

天然CGDDや組換えCGDDをCGDDに特異的な抗体を用いるイムノアフィニティークロマトグラフィーにより実質的に精製する。イムノアフィニティークラムは、CNBr-活性化SEPHAROSE (Amersham Pharmacia Biotech) などの、或る活性化したクロマトグラフィー用レジンと抗CGDD抗体とを共有結合させることにより構成する。結合後に、製造者の使用説明書に従ってこのレジンをブロックし、洗浄する。

【0431】

CGDDを含む培養液をイムノアフィニティークラムに通し、CGDDを優先的に吸着できる条件で (例えば、界面活性剤の存在下において高イオン強度のバッファーで) そのカラムを洗浄する。そのカラムを、抗体とCGDDとの結合を切るような条件で (例えば、或るpH 2~3のバッファー、あるいは高濃度の、例えば尿素またはチオシアン酸イオンなどのカオトロップで) 溶出させCGDDを収集する。

【0432】

1.7 CGDDと相互作用する分子の同定

CGDDまたはその生物活性断片を、¹²⁵Iボルトンハンター試薬で標識する (例えば Bolton, A.E.およびW.M. Hunter (1973) Biochem. J. 133:529-539を参照)。マルチウェルプレートの各ウェルに予め配列しておいた候補の分子群を、標識CGDDと共にインキュベートし、洗浄して、標識されたCGDD複合体を有する全てのウェルをアッセイする。CGDD濃度を様々に変えて得たデータでCGDDの数量、候補分子との親和性および会合についての値を計算する。

【0433】

CGDDと相互作用する分子を別法では、Fields, S.およびO. Song (1989, Nature 340:245-246) に記載の酵母2ハイブリッドシステム (yeast two-hybrid system) や、MATCHMAKER システム (Clontech) などの2ハイブリッドシステムに基づいた市販のキットを用いて分析する。

【0434】

CGDDはまた、ハイスループット型の酵母2ハイブリッドシステムを使用するPATHCALLINGプロセス (CuraGen Corp., New Haven CT) に用いて、遺伝子の2大ライブラリによってコードされるタンパク質間の全ての相互作用を判定できる (Nandabalan, K. 他(2000) 米国特許第6,057,101号)。

【0435】

1.8 CGDD活性の実証

CGDD活性を実証するために、最終分化や細胞周期進行の誘導の測定を、CGDDが哺乳動物細

10

20

30

40

50

胞培養系において生理的に高まったレベルで発現する際に行う。cDNAを高いレベルで発現させる強いプロモーターを持つ哺乳類発現ベクターにcDNAをサブクローニングする。選択されるベクターとしてはPCMV SPORT (Life Technologies, Gaithersburg, MD) およびPCR 3.1 (Invitrogen, Carlsbad, CA) があり、どちらもサイトメガロウイルスプロモーターを持つ。リポソーム製剤あるいは電気穿孔法を用いて、5~10 µgの組換えベクターをヒト細胞株、好適には内皮由来または造血由来の細胞株に、一過的に形質移入する。更に、標識タンパク質をコードする配列を含む1~2 µgのプラスミドを同時に形質移入する。標識タンパク質の発現により、形質移入細胞と非形質移入細胞を区別する手段が与えられる。また、標識タンパク質の発現によって、組換えベクターからのcDNA発現を正確に予想できる。標識タンパク質は、例えば緑色蛍光タンパク質(GFP) (Clontech, Palo Alto, CA)、CD64またはCD64-GFP融合タンパク質から選択できる。フローサイトメトリーは、細胞周期進行か最終分化に先行するか或いは同時に発生する現象を診断する蛍光分子の取込を検出して計量する。このような現象として挙げられるのは、ヨウ化プロピジウムによるDNA染色によって計測される核DNA含量の変化、前方散乱光と90°側方散乱光によって計測される細胞サイズと顆粒性の変化、プロモデオキシウリジン取込量の低下によって計測されるDNA合成の上方または下方調節、特異抗体との反応性によって計測される細胞表面および細胞内におけるタンパク質の発現の変容、および、フルオレセイン抱合したアネキシンVタンパク質の細胞表面への結合によって計測される原形質膜組成の変容がある。フローサイトメトリー法については、Ormerod, M. G. (1994) Flow Cytometry, Oxford, New York, NYに記述がある。

10

20

【0436】

CGDD活性の*in vitro*アッセイは或いは、アンチセンスCGDD RNAを過剰発現する正常なヒト線維芽細胞の形質転換を測定する(Garkavtsev, I. および Riabowol, K. (1997) *Mol. Cell Biol.* 17:2014-2019)。CGDDをコードするcDNAをpLNCXレトロウイルスベクターにサブクローニングし、アンチセンスCGDD RNAの発現を可能にする。結果として得た作成物を、狭宿主性(ecotropic)BOSC23ウイルスパッケージング細胞株に形質移入する。BOSC23培養上澄みを含むウイルスを用いて、両種性CAK8ウイルスパッケージング細胞株を感染させる。CAK8培養上澄みを含むウイルスを用いて、正常なヒト線維芽細胞(Hs68)を感染させる。感染細胞の評価を、次の、転換細胞に特徴的な、定量化できる特性について行う。培地での接触障害の喪失に伴う高密度の成長、懸濁中またはソフトアガーでの成長、コロニーまたは増殖巣の形成、低下した血清要求性、および腫瘍誘発能(免疫不全マウスへの注入時)である。CGDDの活性は、Hs68細胞の形質転換の程度に比例する。

30

【0437】

CGDDの発現は或いは、哺乳動物細胞株で、CGDDをコードする真核生物発現ベクターで細胞を形質転換させて成し得る。真核生物発現ベクターは市販されており、これらベクターを細胞に導入する技術は当業者に周知である。CGDDの細胞局在をアッセイするには、細胞の分画をJiang H. P. 他 (1992; *Proc. Natl. Acad. Sci.* 89: 7856-7860)の記載のように行う。簡潔には、低速遠心法でペレット化した細胞の再懸濁を、バッファ中(10 mM TRIS-HCl, pH 7.4/ 10 mM NaCl/ 3 mM MgCl₂/ 5 mM EDTAに10 µg/mlアプロチニン, 10 µg/mlロイペプチン, 10 µg/mlペプスタチンA, 0.2 mMフッ化フェニルメチルスルホン含有)で行い、ホモジナイズする。ホモジェネートの遠心分離を600×gで5分間おこなう。その上澄みを100,000×gで60分間、超遠心して、粒子画分とサイトゾル画分の分離を行う。核画分を得るため、600×gペレットの、蔗糖溶液での再懸濁(0.25 M蔗糖/ 10 mM TRIS-HCl, pH 7.4/ 2 mM MgCl₂)を行い、再遠心を600×gで行う。各画分からの等量の蛋白をSDS/10%ポリアクリルアミドゲルにかけ、膜にブロットする。ウェスタンブロット分析を行い、これにはCGDD抗血清を用いる。CGDDの局在の評価を、核画分での対応するバンドの、他の画分に比しての強度で行う。CGDDの、細胞画分での存在の試験を或いは、蛍光顕微鏡にCGDD特異的蛍光抗体を用いて行う。

40

【0438】

CGDD活性は、*in vitro*結合アッセイで、その関連するRasスーパーファミリー蛋白と相互

50

作用する能力として実証され得る。候補Rasスーパーファミリー蛋白はグルタチオンSトランスフェラーゼ (GST) との融合タンパク質として発現され、グルタチオン-Sepharose上でのアフィニティークロマトグラフィーによって精製される。100 mM NaCl、2 mM EDTA、5 mM MgCl₂、0.2 mM DTT、100 mM AMP-PNP、および10 mM GDPを含む20mM Trisバッファー (pH 8.0)、30°Cで20分間インキュベーションして、GDPでRasスーパーファミリー蛋白を負荷する。CGDDはバキュロウイルス系でFLAG融合タンパク質として発現される。CGDD-FLAG融合タンパク質を含むこれらのバキュロウイルス細胞の抽出物は、GSTビーズでプレクリアし、次にGST-Rasスーパーファミリー融合タンパク質とインキュベーションする。形成された複合体は、グルタチオン-Sepharoseによって沈殿させ、SDSポリアクリルアミドゲル電気泳動で分離する。分離されたタンパク質はニトロセルロース膜にブロットされ、市販の抗FLAG抗体で探査される。CGDD活性は、複合体中に検出されるCGDD-FLAG融合タンパク質の量に比例する。 10

【0439】

または、LiおよびCohen (Li, L. および S.N. Cohen (1995) Cell 85:319-329)が実証したように、遺伝子の5'端へのアンチセンス配列を設計し、この配列を転写するベクターでNIH 3T3細胞に形質移入して、CGDDの腫瘍化抑制能の測定を成しうる。内因性遺伝子を抑制すると、形質転換した線維芽細胞は、ヌードマウスに導入すれば転移性腫瘍を形成できる細胞塊を産生しうる。

【0440】

CGDD活性のアッセイは或いは、注入CGDDの、母性転写物の分解への効果を測定する。卵母細胞の、Swissアルビノマウスからの収集、注入、培養の手順は、Stutz(前出)に記載がある。母性RNAの分解の、対照卵母細胞に比しての低下が、CGDD活性を示す。CGDD活性の測定は或いは、精製CGDDがRNAアーゼに結合する能力として、実施例 17に記載のアッセイで測定するように成される。 20

【0441】

CGDD活性のアッセイは或いは、CGDD発現プラスミドで形質移入したCOS細胞におけるシンシウム形成を、2成分融合(two-component fusion)アッセイ(前出Miに記載)で測定する。このアッセイが利用する事実は、ヒトのインターロイキン12 (IL-12)がヘテロダイマーであり、その構成サブユニットの分子量が35 kD (p35) と40 kD (p40)であることである。p35の遺伝子を持つ発現プラスミドで形質移入したCOS細胞を、p40とCGDDとの遺伝子を持つ発現プラスミドを同時移入したCOS細胞と混ぜる。IL-12活性の、結果として得た馴化培地でのレベルが、このアッセイのCGDD活性に相当する。合胞体形成はまた、顕微鏡で測定しうる(Mi他、前出)。 30

【0442】

CGDD活性の別のアッセイは、スイスマウス3T3細胞における新規開始DNA合成の量として細胞増殖を測定する。CGDDをコードするポリヌクレオチドを含むプラスミドを当分野で公知の方法を用いて、静止状態の3T3培養細胞に移入する。一過性に形質移入された細胞は次に、[³H]チミジンまたは放射性DNA前駆体([³²P]ATP等)の存在下でインキュベートされる。適用可能な場合には、様々な量のCGDDリガンドが、形質移入された細胞に加えられる。酸によって沈殿され得るDNAへの[³H]チミジンの取込は適切な時間間隔で測定され、その取込量は新規に合成されたDNAの量とCGDD活性に正比例する。 40

【0443】

CGDD活性を、或いはサイクリン-ユビキチンライゲーションアッセイにより測定する(Townley, F.M. 他(1997) Proc. Natl. Acad. Sci. USA 94:2362-2367)。反応液は体積10 μl中に、40 mMのTris.HCl (pH 7.6)、5 mMのMg Cl₂、0.5 mMのATP、10 mMのホスホクレアチン、50 μgのクレアチン・ホスホキナーゼ/ml、1 mg還元カルボキシメチル化ウシ血清アルブミン/ml、50 μM ユビキチン、1 μMのユビキチンアルデヒド、1~2 pmol ¹²⁵I-標識化サイクリンB、1 pmol E1、1 μMのオカダ酸、M期分画1Aのタンパク質10 μg (活性E3-Cを含み原則的にE2-C不在)と可変量CGDDを含む。反応液を18°Cで60分間インキュベートする。次にSDSポリアクリルアミドゲル上で電気泳動法によりサンプルを分離する。形成 50

された¹²⁵I-サイクリン-ユビキチンの量をPHOSPHORIMAGER分析で定量する。サイクリン-ユビキチン形成の量は反応物中のCGDDの活性に比例する。

【0444】

CGDD活性のアッセイは或いは、放射標識ヌクレオチド、例えば[α³²P]ATPを用い、放射ラベルの、DNA合成時のDNAへの取込か、アポトーシスに伴うDNAの断片化を測定する。CGDDをコードするcDNAを持つプラスミドで、当分野で周知の方法を用いて哺乳動物細胞に形質移入する。次に細胞のインキュベートを、放射標識ヌクレオチドと共に、種々の時間長で行う。染色体DNAを収集し、放射能の検出にシンチレーションカウンターを用いる。放射ラベルの、染色体DNAへの取込は、細胞周期の刺激の程度に比例する。CGDDのアポトーシス促進を判定するには、染色体DNAを上記のように収集し、分析にポリアクリルアミドゲル電気泳動を用い、当分野で周知の方法で行う。DNA断片化は定量のため非移入対照細胞と比較し、これがCGDDのアポトーシス活性に比例する。

10

【0445】

CGDDのシクロフィリン活性の別法での測定には、キモトリプシン共役アッセイを用い、シス〜トランス相互転換の速度を測定する(Fischer, G., Bang, H., および Mech, C. (1984) Biomed. Biochim. Acta 43: 1101-1111)。キモトリプシンを用いてXaa-Proペプチド結合でのトランス基質切断活性を推計する。ここでシス〜トランス異性化の速度定数を得るには、緩徐相での基質水解の速度定数を測定する。サンプルのインキュベートを免疫抑制剤CsAかFK506の有無の中で行い、反応開始にはキモトリプシンを添加し、蛍光反応を測定する。酵素的速度定数を求める式は $k_{app} = k_{H2O} + k_{enz}$ であり、ここでは一次速度式が示され、1単位のPPIase活性の定義は $k_{enz} \text{ (s}^{-1}\text{)}$ である。

20

【0446】

蛍光モニタリングアッセイで活性化Ras検出にRRP22を用いるには次のようにする。c-Raf1(減数分裂の再開時に活性化するキナーゼ)のRRP22結合ドメイン(RRP22BD)の合成は、2つの無保護ペプチドセグメントから、ネイティブ化学ライゲーションで成される。ニトロベンズ-2-オキサ-1,3-ジアゾールとクマリル・クロモフォア(coumaryl chromophores)に基づく構造の2つの蛍光アミノ酸がRRP22BD/RRP22-GTP結合面の近くへ取込まれた後、His(6)からなるC末端タグの導入が起きる。この部位特異的に修飾された蛋白の、Ras-GTPへの結合の K_D 値を、野生型RBDと比較する。Ras-GTPの検出を、100 nMのレンジ内で、C末端His(6)タグ修飾した蛍光RBDの、Ni-NTAコートした表面への固定化で行う。Ras-GDPは固定化RBDには結合しないので、不活性および活性化Rasを区別できる(Becker, C. F. (2001) Chem. Biol. 8:243-252)。

30

【0447】

CGDDのRas結合アッセイの方法は、Vavvas他(前出)に見られる。CGDD発現はGST融合タンパク質として成され、GST-CGDD融合体のインキュベーションをRasと共に、GTPγSかGDPβSの存在下で行う。グルタチオン-SEPHAROSEビーズ(APB)を添加し、GST-CGDD融合体とGST-CGDD-Ras複合体を溶液から回収する。蛋白質をグルタチオン-セファロースのビーズからSDSサンプル緩衝液で溶出し、SDS-PAGEで分離する。電気泳動後、タンパク質をPVDF膜(APB)に移し、Rasのプローブをモノクローナル抗Ras抗体で行う。

40

【0448】

Wnt5aの、細胞間接触による調節は、蛋白チロシンキナーゼを選択的にブロックする種々の代謝剤(ゲニステイン)やサイトカラシンDを、HB2(正常な乳腺上皮細胞株)に添加して見られる。サイトカラシンD処理後の細胞骨格再編成によってWnt5aの誘導が起きる。これは、細胞形態の変化を伴う。サイトカラシンD処理した癌細胞株は、細胞形態の変化もWnt5a誘導も見せない(Jonsson, M. 他(1998) Br. J. Cancer 78:430-438)。

【0449】

19 CGDD結合アッセイ

CGDDシクロフィリンの或る定量イムノアッセイは、その、免疫抑制剤シクロスポリンへの立体特異的結合の親和性を測定する(Quesniaux, V.F., 他(1987) Eur. J. Immunol. 17: 1359-1365)。このアッセイでは、シクロフィリン-シクロスポリン複合体を固相にコート

50

し、結合の検出に、抗シクロフィリンウサギ抗血清を抗グロブリン-酵素抱合体で増強して用いる。CGDDイムノフィリン、シクロフィリンの、免疫抑制剤シクロスポリンとの、決定的残基での複合体形成は、免疫抑制作用(例えば腫瘍化時に発生)を促進する。

【0450】

FKBPと免疫抑制剤との気相での非共有結合測定のために開発された或る結合アッセイは、エレクトロスプレーイオン化質量分析を用いる(Trepanier, D.J., 他(1999) Ther. Drug Monit. 21: 274-280)。エレクトロスプレーイオン化でイオンを発生するには、高電荷の液滴の極細スプレーを強い電場の存在下で発生させる。液滴のサイズが小さいほど、表面の電荷密度が高い。イオンは静電的に質量分析器に誘導され、ここで反対電荷のイオンの発生が、空間的に分離したソースで成された後、毛細管注入口へ掃引され、ここでイオン流が合流され、反応が起きる。結合と無結合のFKBP/免疫抑制剤複合体の荷電状態を比較して、相対結合親和性を確立でき、in vitroの結合および免疫抑制作用と相関させうる。

10

【0451】

当業者には、本発明の要旨および精神から逸脱しない範囲での、記載した本発明の方法およびシステムの、種々の修正および変更の手段は自明であろう。本発明について説明するにあたり幾つかの実施例に関連して説明を行ったが、本発明の請求の範囲が、そのような特定の実施例に不当に制限されるべきではないことを理解されたい。分子生物学または関連分野の専門家には明らかな、本明細書に記載する本発明の実施方法の様々な修正は、明確に特許請求の範囲内にあるものとする。

【0452】

(表の簡単な説明)

表1は、本発明の完全長ポリヌクレオチド配列およびポリペプチド配列の命名法の概略を示す。

20

【0453】

表2は、GenBank識別番号と、本発明のポリペプチドに最も近いGenBank相同体の注釈とを示す。また、各ポリペプチドとその相同体(1つ以上)が一致する確率スコアも併せて示す。

【0454】

表3は、予測されるモチーフおよびドメインなど、本発明のポリヌクレオチド配列の構造的特徴を、ポリペプチドの分析に用いる方法、アルゴリズムおよび検索可能なデータベースと共に示す。

30

【0455】

表4は、本発明のポリヌクレオチド配列を構築するために用いたcDNA断片やゲノムDNA断片を、ポリヌクレオチド配列の選択した断片と共に示す。

【0456】

表5は、本発明のポリヌクレオチドの代表的なcDNAライブラリを示す。

【0457】

表6は、表5に示したcDNAライブラリの作製に用いた組織およびベクターを説明する付表である。

【0458】

表7は、本発明のポリヌクレオチドとポリペプチドの分析に用いたツール、プログラム、アルゴリズムを、適用可能な説明、参照文献および閾値パラメータと共に示す。

40

【0459】

表8は、本発明のポリヌクレオチド配列に見られる一塩基多型を、種々のヒト集団での対立遺伝子(アレル)頻度と共に示す。

【0460】

【表1-1】

表 1 - 1

Incyte プロジェクト ID	ポリペプチド SEQ ID NO:	Incyte ポリペプチド ID	ポリヌクレオチド SEQ ID NO:	Incyte ポリヌクレオチド ID	CA2 試薬
1351608	1	1351608CD1	22	1351608CB1	
4259314	2	4259314CD1	23	4259314CB1	2747431CA2, 3421573CA2
3660046	3	3660046CD1	24	3660046CB1	
3016416	4	3016416CD1	25	3016416CB1	7291179CA2
2133755	5	2133755CD1	26	2133755CB1	90067312CA2
5259957	6	5259957CD1	27	5259957CB1	5259957CA2
55029783	7	55029783CD1	28	55029783CB1	4949559CA2
8032202	8	8032202CD1	29	8032202CB1	
6937367	9	6937367CD1	30	6937367CB1	
3876510	10	3876510CD1	31	3876510CB1	
4900076	11	4900076CD1	32	4900076CB1	
1543848	12	1543848CD1	33	1543848CB1	
6254070	13	6254070CD1	34	6254070CB1	7677372CA2, 90072067CA2, 90072183CA2, 90072191CA2
1289839	14	1289839CD1	35	1289839CB1	8698064CA2
5565648	15	5565648CD1	36	5565648CB1	
2764456	16	2764456CD1	37	2764456CB1	
5734806	17	5734806CD1	38	5734806CB1	
7495168	18	7495168CD1	39	7495168CB1	
7483131	19	7483131CD1	40	7483131CB1	
4558650	20	4558650CD1	41	4558650CB1	

表1-2

Incyte プロジェクト ID	ポリペプチド SEQ ID NO:	Incyte ポリペプチド ID	ポリヌクレオチド SEQ ID NO:	Incyte ポリヌクレオチド ID	CA2 試薬
7506195	21	7506195CD1	42	7506195CB1	

10

20

30

40

表 2-1

ポリペプチド SEQ ID NO:	Incyte ポリ ペプチド ID	GenBank ID NO: または PROTEOME ID NO:	確率スコア	注釈
1	1351608CD1	g3170887	0.0	[マウス] ユビキチン蛋白質リガーゼ E3 α (Kwon, Y.T., 他 (1998) The mouse and human genes encoding the recognition component of the N-end rule pathway) Proc. Natl. Acad. Sci. U.S.A. 95: 7898-7903.)
2	4259314CD1	g11990422	8.2E-198	[ヒト] MOP-4
3	3660046CD1	g12656270	0.0	[マウス] ユビキチン蛋白質リガーゼ Nedd4-2. Kamynina, E. 他 (2001) A novel mouse Nedd4 protein suppresses the activity of the epithelial Na ⁺ channel. FASEB J. 15:204-214.
4	3016416CD1	g4868435	3.1E-45	[ヒト] アポトーシス関連タンパク質 APR-2
5	2133755CD1	g12248402	8.4E-281	[ヒト] シスプラチン耐性関連タンパク質 CRR9p Yamamoto, K. 他 (2001) Biochem. Biophys. Res. Commun. 280:1148- 1154.
6	5259957CD1	g1903384	3.7E-120	[ヒト] 優先的に発現される黒色腫抗原 Ikeda, H., 他 (1997) Characterization of an antigen that is recognized on a melanoma showing partial HLA loss by CTL expressing an NK inhibitory receptor. Immunity 6, 199-208.
7	55029783CD 1	g8131896	1.7E-188	[ヒト] 膀胱癌過剰発現タンパク質
8	8032202CD1	g9081879	4.3E-11	[ヒト] 上咽頭癌関連タンパク質 1 He, Z.W., 他 (2000) Cloning of a novel gene associated with human nasopharyngeal carcinoma. Chin. Sci. Bull. 45, 2267-2272.
9	6937367CD1	g8650435	1.5E-299	[マウス] RalGDS 様タンパク 3. Shao, H. and Andres, D.A. (2000) a novel RalGEF-like protein, RGL3, as a candidate effector for rit and Ras. J. Biol. Chem. 275, 26914-26924.

10

20

30

40

表 2-2

ポリペプチド SEQ ID NO:	Incyte ポリ ペプチド ID	GenBank ID NO: または PROTEOME ID NO:	確率スコア	注釈
10	3876510CD1	g5870832	7.3E-259	[マウス] skm-BOPI. Hwang, I. and Gottlieb, P.D. (1997) The Bop gene adjacent to the mouse CD8b gene encodes distinct zinc-finger proteins expressed in CTLs and in muscle. J. Immunol. 158, 1165-1174.
11	4900076CD1	g9392657	2.2E-158	[ヒト] PR ドメイン含有タンパク 9. Jiang, G.L., and Huang, S. (2000) The yin-yang of PR-domain family genes in tumorigenesis. Histol. Histopathol. 15, 109-117.
12	1543848CD1	g4809026	3.2E-121	[ヒト] skp1 相同体の G2 アレルの抑制因子; Sgt1. Kitagawa, K., 他 (1999) Mol. Cell 4, 21-33.
13	6254070CD1	g348918	2.5E-205	[ヒト] hWNT5A. Clark, C. C. 他 (1993) Molecular cloning of the human proto-oncogene Wnt-5A and mapping of the gene (WNT5A) to chromosome 3p14-p21. Genomics 18:249-260.
14	1289839CD1	g1666073	1.7E-48	[ヒト] RRP22 タンパク. Zucman-Rossi, J. 他 (1996) Identification of new members of the Gas2 and Ras families in the 22q12 chromosome region. Genomics 38:247-254.
15	5565648CD1	g6164747	5.3E-36	[ヒト] F ボックス 蛋白 Fbx22. (Cenciarelli, C. et al. (1999) Identification of a family of human F-box proteins. Curr. Biol. 9: 1177-1179.)
16	2764456CD1	g6630609	3.6E-252	[ヒト] サイクリン E 結合タンパク 1. (Mitsui, K. et al. (1999) A novel human gene encoding HECT domain and RCC1-like repeats interacts with cyclins and is potentially regulated by the tumor suppressor proteins. Biochem. Biophys. Res. Commun. 266:115-122.)

10

20

30

40

表 2 - 3

ポリペプチド SEQ ID NO:	Incyte ポリ ペプチド ID	GenBank ID NO: または PROTEOME ID NO:	確率スコア	注釈
17	5734806CD1	g5923891	0.0	[ヒト] シクロフィリン関連タンパク Anderson, S.K. et al. (1993) Proc. Natl. Acad. Sci. U.S.A. 90:542-546.
18	7495168CD1	g13277570	6.0E-108	[ヒト] (BC003694) 腫瘍壊死因子、 α 誘導タンパク 1 (内皮)
19	7483131CD1	g2330015	8.3E-21	[ヒト] アポトーシス性プロテアーゼ活性化因子 1 (Apaf-1) Zou, H. et al. (1997) Cell 90, 405-413
20	4558650CD1	g4416181	2.4E-234	[マウス] ES18 Park, E.J. et al. (1999) Nucleic Acids Res. 27, 1524-1530
21	7506195CD1	g4416181	5.2E-155	[マウス] ES18 Park, E.J. et al. (1999) <i>supra</i>

表 3 - 1

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
1	I351608CD17381	S16 S100 S277 S570 S586 S625 S689 S715 S816 S833 S989 S993 S1053 S1078 S1257 S1272 S1278 S1489 S1582 S1605 T61 T162 T194 T250 T433 T480 T727 T752 T757 T787 T798 T882 T973 T1036 T1080 T1512 T1549 Y99 Y216 Y677	N281 N711 N750 N785 N1043	推定 Zn フィンガー、N-recognin 内 (N 末端則経路の認識成分): Q86-E156	HMMER_PFAM	
					膜貫通ドメイン: V633-E651, F864-R881, S1278-L1295, I1378-Y1406, S1421-L1449, G1483-Y1507, G1640-R1660	TMAP
					ユビキチンタンパクリガーゼ E3 COMPONENT NRECOGNIN ユビキチンタンパクリガーゼ E3 α リガーゼ PD149323: L1202-Y1485	BLAST_PRODROM

表 3 - 2

SEQ ID NO.	Incyle ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
1					リガーゼ ユビキチン蛋白 E3 成分 N 末端認識 N RECOGNIN ユビキチン抱合 PD006492: A355-E897, T752-I1225, I341-L535, F24-S100	BLAST_PRODOM
1 (続)					ユビキチン蛋白リガーゼ E3 COMPONENT NRECOGNIN ユビキチン蛋白リガーゼ E3 α リガーゼ PD149312: M1-F85	BLAST_PRODOM
					リガーゼ ユビキチン蛋白 E3 成分 N 末端認識タンパク NRECOGNIN ユビキチン抱合 PD010181: P37-Q284, P37-L329	BLAST_PRODOM
					ロイシンジッパーパターン: L593-L614	MOTIFS
2	4259314CD1	389	S108 S149 S272 S317 S340 S360 T25 T89 T93 T208 T209 T230 T246 T277 T293 T364 Y132	N143 N220 N257	Thif ファミリー: A60-G200	HMMER_PFAM

表 3－3

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
					膜貫通ドメイン： G70-K86 N 末端はサイトソル内れない。	TMAP
					ユビキチン活性化酵素 E1 ユビキチン抱合リガーゼリピート 多重遺伝子家族タンパク PD005434; C197-E363	BLAST_PRODOM
2 (続)					タンパク質 酵素 ユビキチン活性化ユビキチン E1 抱合 リガーゼリピート 合成 多重遺伝子 PD000731; S62-F196	BLAST_PRODOM
					ユビキチン活性化酵素 DM00412 P22314 44-293; A35-Q290 DM00412 P20973 37-288; E38-I287 DM00412 P31252 40-291; S33-I287 DM00412 P22515 8-256; T34-I287	BLAST_DOMO

表 3 - 4

SEQ ID NO.	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
3	3660046CD1	854	S16 S104 S110 S130 S182 S191 S200 S214 S221 S268 S366 S510 S565 S638 S771 S832 T6 T154 T248 T402 T403 T437 T453 T505 T624 T648 T679 T784 Y688	N59 N95 N287 N569	HECT-ドメイン (ユビキチン-トランスフェラーゼ) : F548-V853	HMMER_PFAM
					WW ドメイン: L266-P295, L378-P407, L74-P103, L429-P458	HMMER_PFAM
					膜貫通ドメイン: F589-F610	TMAP
3 (続)					WW/isp5/WWP ドメイン蛋白質: BL01159: Y89-P103	BLIMPS_BLOCKS
					WW ドメインシグネチャ: PR00403: L378-R391, Y281-P295	BLIMPS_PRINTS
					HECT-ドメイン (ユビキチン-トランスフェラーゼ): PF00632: F711-D717, W761-P788, L815-N846	BLIMPS_PFAM

表 3 - 5

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
					タンパク質リガーゼユビキチン抱合リビートユビキチン蛋白 DNA 結合 推定 線原性 PD002225; I623-F850, A543-L669	BLAST_PRODUM
					NEDD4 タンパク質 PD092083; S104-G265 PD064436; I296-F377 PD033209; M1-S63	BLAST_PRODUM
					HECT ドメイン DM01690 P39940 S13-808; N556-F850 DM01690 P51593 P-306; N556-F850 DM01690 A38919 785-1082; F555-D851 DM01690 P53119 615-909; N556-D851	BLAST_DOMO
					WW/isp5/WWP ドメインシグネチャ; W78-P103, W270-P295, W382-P407, W433-P458	MOTIFS
4	3016416CD111		S69 T88		signal_cleavage: M1-G26	SPSCAN
5	2133755CD538		S3 S84 S161 S219 S231 S268 S317 S322 S387 S389 S491 S526 T60 T160 T385 T470 T476	N91 N101 N229	シグナルペプチド: M1-T23	HMMER

表 3 - 6

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
					膜貫通ドメイン : M1-Y28 T282-S310 T323-V349 Y429-K457 P465-R493 N 末端はサイトソル内。	TMAP
					タンパク質 胸腺上皮細胞表面抗原 R166.2 T13H5.2 PD025254; E87-P410	BLAST_PRODOM
					タンパク質 R166.2 T13H5.2 PD042934; L411-K530	BLAST_PRODOM
6	5259957CD1	474	S2 S26 S45 S77 S80 S122 S128 S192 S277 S305 S317 T27 T51 T139 T157 T439	N155 N270	ロイシンリッチリピート : Q320-L343, T349-S373, K99-P124	HMMER_PFAM
					メラノーマ優先的発現抗原 KIAA0014 タンパク質 DJ845O24.3 PRAME LIKE OF PD043129; L69-C472, A5-E125	BLAST_PRODOM
7	55029783CD1	354	S12 S19 S168 S186 S341 T7 T15	N84 N133 N138 N279	膜貫通ドメイン : T73-L99 S116-S135 I176-V204 Y231-F251 W260-L280 T289-G309 S318-H337 N 末端はサイトソル内	TMAP
7 (続)					細胞接着配列: R39-D41	MOTIFS

10

20

30

40

表 3 - 7

SEQ ID NO.	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
8	8032202CD1	272	S82 S226 T237 T256	N36 N67 N95 N118	signal_cleavage: M1-A20	SPSCAN
					シグナルペプチド: M2-A20, M2-K23, M1-P25, M1-A32	HMMER
					膜貫通ドメイン: W6-R34 N-末端はサイトゾル内にならない	TMAP
9	6937367CD1	710	S30 S36 S40 S170 S221 S247 S277 S342 S377 S388 S402 S495 S512 S520 S526 S540 S559 S580 S591 S601 T63 T99 T256 T290 T372 T522 T635	N339	Ras 会合(RalGDS/AF-6) ドメイン: E613-G700	HMMER_PFAM
					RasGEF ドメイン: L244-K454	HMMER_PFAM
					タンパク質 因子 放出 ガニンヌクレオチド SH3 ドメインのガニンヌクレオチド SON PD002030: L243-W456	BLAST_PRODOR
					LTE1 タンパク質 DM08605[A56234]-495: L11-S499	BLAST_DOMO

表 3 - 8

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
9 (続)					デアニンヌクレオチド解離抑制因子 CDC25ファミリー DM01741 P04821 I062-I534; L69-L393, V421-L498	BLAST_DOMO
10	3876510CD1	490	S128 S243 S251 T183 T444 Y325 Y458	N211 N257	MYND Zn フィンガー : C52-C90	HMMER_PFAM
					膜貫通ドメイン: I348-R376 N-末端はサイトソル内	TMAP
					Zn フィンガー PF00642: L62-H72	BLIMPS_PFAM
					タンパク質 Zn フィンガー BOP 核 DNA 結合選択的スプライシング 金属結合 PD034977: E327-Q490 PD129736: E7-V51 PD129734: Q280-H326	BLAST_PRODOM
					タンパク質 Zn フィンガー (1 証拠 (IEVIDENCE) = 予想されたマッチによる) 推定 BOP 様 ORF YPL165C 亜鉛 PD005436: A92-L287	BLAST_PRODOM

表 3 - 9

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
11	4900076CD1	599	S80 S100 S181 S203 S277 S307 S333 S363 S389 S475 S503 S559 S587 T28 T33 T63 T72 T189 T235 T318 T346 T395 T430 Y325 Y437	N44 N311	KRAB ボックス: V62-S112	HMMER_PFAM
					Zn フィンガー、C2H2 タイプ: Y437-H459, Y521-H543, Y325-H347, Y297-H319, Y577-H599, Y409-H431, Y549-H571, Y269-H291, Y465-H487, Y381-H403, Y493-H515, Y353-H375	HMMER_PFAM
					Zn フィンガー、C2H2 タイプ BL00028: C299-H315	BLIMPS_PRINTS
					タンパク質 Zn フィンガー 亜鉛 PD01066: F64-A102	BLIMPS_PRODOM
					Zn フィンガー-蛋白質 PD017719: P352-F586 PD101192: A102-N242 PD000072: K295-C358	BLAST_PRODOM
					Zn フィンガー、C2H2 タイプ、ドメイン DM00002[P08042]314-358: C358-H403, C526-H571, C442-H487	BLAST_DOMO

表 3-10

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
11 (続)					Zn フィンガー、C2H2 タイプ、ドメイン: C271-H291 C299-H319 C327-H347 C355-H375 C383-H403 C411-H431 C439-H459 C467-H487 C495-H515 C523-H543 C551-H571 C579-H599	MOTIFS
12	1543848CD1	365	S11 S17 S161 S202 S214 S239 S331 T101 T126 T141 T234 T240 T265 Y47	N77 N78 N146 N159 N329 N343	signal_cleavage: M1-A68	SPSCAN
					TPR ドメイン: A45-N78, S79-T112	HMMER, PFAM
					タンパク質 D1054.3 SGT1 PD022846: K194-Y365	BLAST, PRODOM
13	6254070CD1	365	S7 S101 S198 S233 T280 T337 T356	N99 N105 N297 N311	signal_cleavage: M1-A22	SPSCAN
					シグナルペプチド: M6-V25, M1-S30, M6-S30	HMMER
					wnt ファミリー: L53-K365	HMMER, PFAM
					膜貫通ドメイン: S8-W32 N-末端はサイトソル内にはない	TMAP
					Wnt-1 ファミリー蛋白: BL00246: G81-C100, G116-S150, W162-E186, M203-Y255, G319-C364	BLIMPS_BLOCKS
					Wnt-1 ファミリーシグネチャ wnt1.prf: M203-K252	PROFILESAN

表 3-11

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
					タンパク質 発生の 糖タンパク質 前駆体シグナル WNT1 WNT5A WNT2 細胞外マトリックス PD000810: D161-K365, L53-G267	BLAST_PRODOM
					WNT-1 ファミリー DM00403 Q06442 13-358: Q23-K365	BLAST_DOMO
13 (続)					WNT-1 ファミリー DM00403 P09544 26-348: L53-K365	BLAST_DOMO
					WNT-1 ファミリー DM00403 P47793 24-351: Q41-K365	BLAST_DOMO
					WNT-1 ファミリー DM00403 P34889 25-359: S30-K365	BLAST_DOMO
					Wnt-1 ファミリーシグネチャ C223-C232	MOTIFS
14	1289839CD1	203	S159 T4 T37 T70 T105 T150	N142	Ras ファミリー: R6-M203	HMMER_PFAM
					膜貫通ドメイン: C79-R107, N-末端はサイトソール内がない	TMAP
					形質転換性タンパク質 P21 RAS シグネチャ PR00449: S119-L132, Y155-V177, Y5-Y26, V46-V68	BLIMPS_PRINTS
					RAS 形質転換性タンパク質 DM00006 S41960 3-148: R6-E157	BLAST_DOMO
					RAS 形質転換性タンパク質 DM00006 J55401 3-148: R6-E157	BLAST_DOMO
					RAS 形質転換性タンパク質 DM00006 P48555 8-153: Y5-E157	BLAST_DOMO

表 3-12

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
					RAS 形質転換性タンパク質 DM00006P11234[11-157: V2-E157]	BLAST_DOMO
					ATP/GTP 結合部位モチーフ A (P ループ) G11-S18	MOTIFS
15	5565648CD1	403	S13 S113 S128 S162 S263 S401 T62 T127 T242	N198 N257	F-ボックスドメイン: S19-T67	HMMER_PFAM
15 (続)					膜貫通ドメイン: A344-G361 N-末端はサイトソル内。	TMAP
16	2764456CD1	1022	S46 S216 S224 S242 S290 S339 S368 S385 S397 S647 S669 S690 S793 S801 S807 S867 S915 S975 S995 T17 T118 T124 T259 T319 T323 T362 T375 T398 T427 T431 T443 T448 T459 T550 T552 T729 T850 T904	N263 N865	HECT-ドメイン (ユビキチン-トランスフェラーゼ): E726-V1015	HMMER_PFAM

表 3 - 1 3

SEQ ID NO.	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
			T939 T951 T973 Y758			
					染色体凝縮の調節因子 (RCC1) : K93-K145, D254-H304, D146-L198, G200-Q253, R43-H92	HMMER_PFAM
					膜貫通ドメイン : R757-K785 N-末端はサイトソール内でない。	TMAP
					染色体凝縮の調節因子 (RCC1) タンパク質 BL00625: G136-K164, V238-S271	BLIMPS_BLOCKS

表 3-14

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
16 (続)					染色体凝縮の調節因子 (RCC1) シグネチャ: rcc1_1.prf: V186-V238 rcc1_2.prf: A60-F112, G110-E165, P221-G269	PROFLESCAN
					染色体凝縮調節因子 RCC1 シグネチャ: PR00633: R9S-E111, V149-E165, I241-V257	BLIMPS_PRINTS
					HECT-ドメイン (ユビキチン-トランスフェラーゼ) PF00632: V836-G842, L930-H957, H979-N1010	BLIMPS_PFAM
					タンパク質リガーゼ ユビキチン抱合リビート ユビキチン蛋白 DNA 結合 推定 癌原性 PD002225: V716-F1014	BLAST_PRODOM
					ユビキチン抱合 PD136613: V348-L674, P68-G110, P119-G163, E11-Q70, P227-G307, Q169-L228 PD071828: K164-L198	BLAST_PRODOM

表 3 - 1 5

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
					タンパク質 リピート グアニンスクレオチド放出因子 調節因子 細胞周期 有糸分裂 PD001424; K93-G163	BLAST_PRODOR
16 (続)					HECT ドメイン DM01690 P39940 513-808; F721-F1014 DM01690 P40985 578-891; P731-D905, V836-F1014, Q212-K226 DM01690 A38919 785-1082; Y758-F1014 DM01690 P53119 615-909; K755-F1014	BLAST_DOMO
					染色体凝縮の調節因子(RCC1) シグネチャ 2: A28-L38, V80-V90, V133-L143, V186-L196, I241-L251	MOTIFS

表 3 - 1 6

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
17	5734806CD1	1462	S145 S180 S206 S210 S212 S214 S221 S235 S243 S244 S260 S274 S328 S348 S350 S352 S383 S393 S394 S402 S462 S466 S470 S476 S483 S494 S495 S500 S502 S513 S517 S531 S553 S563 S565 S670 S681 S688 S695 S722 S747 S761 S766 S768 S773 S774 S776 S798 S802	N119 N203 N863 N870 N900 N937 N1047	シクロフィリン型ペプチジルプロリルシステラントランス異性化酵素: Q8-A177	HMIMER_PFAM

表 3-17

SEQ ID NO:	Incyclic Peptide ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
17 (続)			S808 S820 S846 S848 S855 S882 S888 S890 S898 S906 S915 S917 S939 S949 S955 S957 S964 S1044 S1060 S1061 S1066 S1076 S1180 S1202 S1243 S1275 S1295 S1296 S1303 S1318 S1346 S1372 S1375 S1378 S1382 S1390 S1398 S1417 S1435 S1441 S1443		膜貫通ドメイン：P131-I150 N-末端はサイトゾル内にはない	TMAP
			S1445 S1454 S1458 T35 T53 T126 T193 T337 T419 T453 T471 T859 T931 T989 T1049 T1057 T1082 T1207 T1252 T1283 Y90 Y1461		シクロフィリン型ペプチジルプロロリルシートランスマシ性化酵素シグネチャ BL00170: Q21-K47, Y59-N98, A106-I150	BLIMPS_BLOCKS

表 3 - 1 8

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
					シクロフィリン型ペプチジルプロリルシス-トランス異性化酵素シグネチャおよびプロファイル: D30-D96	PROFILES CAN
17 (続)					シクロフィリンペプチジルプロリルシス-トランス異性化酵素シグネチャ PR00153: L27-L42, F64-G76, F107-Q122, Q122-D134, G135-I150	BLIMPS_PRINTS
					NK 腫瘍認識タンパク ナチュラルキラー細胞 シクロフィリン関連 NKTR シクロロスポリリン異性化酵素 ROTAMASE リピート膜貫通 PD145736: N729-T1328 PD078637: G557-Y680 PD055818: C447-G530	BLAST_PROD OM
					異性化酵素 ROTAMASE シクロフィリンシス-トランスペプチジルプロリル PPI アーゼ シクロロスポリリン多重伝子族タンパク PD000341: F11-L176	BLAST_PROD OM
					認識; 腫瘍; プロリル; 天然; DM08077 IP30414 230-1403: L176-R1348 IB47328 230-1507: L176-Y1461	BLAST_DOM O
					シクロフィリン型ペプチジルプロリルシス-トランス異性化酵素 DM00129 IB47328 58-229: Q4-V175 IP30414 58-229: Q4-V175	BLAST_DOM O

表 3 - 1 9

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
18	7495168CD1	329	S22 S71 S89 S110 S155 S186 T55 T60 T143 T166T187 T245 T250 T271 T315	N176	シクロフィリン型ペプチジルプロリルシス-トランス異性化酵素シグネチャ: Y59, G76 K+チャネル四量体化ドメイン: K41-Q138	MOTIFS HMMER_PFAM
19	7483131CD1	476	S38 S128 S136 S159 S175 S194 S333 S370 S439 S445 T7 T33 T164 T184 T268 T444 T462	N150 N365 N460	タンパク質 EDP1 腫瘍壊死因子 α 誘導性 内皮型 B12 PD037429: L118-D329 signal_cleavage: M1-L30	BLAST_PRODOR SPSCAN
					SAM ドメイン (不妊 α モチーフ): E330-R394 WD ドメイン, G- β リピート: L4-S38, I273-Q309, Y231-D267, P46-N82, A90-N125, C172-I218, L131-D167	HMMER_PFAM
					Trp-Asp (WD-40) リピートシグネチャ: L243-A290	PROFILESAN

表 3 - 2 0

SEQ ID No:	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
					β グタンパク質(トランスデュシン)シグネチャ PR00319:L235-G251, L254-T268, I273-A288, P291-W308 Gタンパク質 β WD-40 リピートシグネチャ PR00320: L69-T83, L296-F310	BLIMPS_PRINTS
					Trp-Asp (WD) リピートシグネチャ: L69-T83, L112-A126 signal_cleavage: M1-A68	MOTIFS
20	4558650CD1	485	S119 S170 S343 S363 T153 T312		タンパク質 リピートシグナル前駆体ブリアン糖タンパク質 核内 GPI アンカー 脳主 要 PD001091:G10-P123 プロテインチロシンホスファターゼ TD14 EC 3.1.3.48 加水分解酵素 PD180360: P4-R187, E323-K357	SPSCAN
20 (続)						BLAST_PRODROM
					FORMIN; DM04565[Q05858]1-1212:P12-G115 線維性コラーゲン カルボキシル末端 DM00042[A41132]43-133: P12-C95 DM00042[S21930]37-137:P12-F98 プロリンリッチタンパク質 DM03894[A39066]1-159: P14-P129 signal_cleavage: M1-A68	BLAST_DOMO
21	7506195CD1	406	S119 S170 T153 T312			SPSCAN

10

20

30

40

表 3-21

SEQ ID NO:	Incye ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
					タンパク質 リピート シグナル 前駆体 プリオン糖タンパク質 核内 GPI アンカー 脳主要 PD001091: G10-P123, R11-P129	BLAST_PRODOR
					線維性コラーゲン カルボキシル末端 DM00042 A41132 43-133: P12-C95, P14-P117	BLAST_DOMO
					線維性コラーゲン カルボキシル末端 DM00042 S21930 37-137: P12-F98, P20-P120	BLAST_DOMO
					プロリンリッチタンパク質 DM03894 A39066 1-159: P14-P129	BLAST_DOMO
					フォルミン (FORMIN) ; DM04565 Q05858 1-1212: P12-G115	BLAST_DOMO

表 4-1

ポリヌクレオチド SEQ ID NO/ Incyte ID/配列長	配列断片
22/1351608CB1 / 7742	1-644, 1-733, 637-5226, 3489-3638, 3563-3836, 3661-4146, 4346-4436, 4349-4563, 4457-4914, 4545-4994, 4559-4972, 4574-4755, 4635-4889, 4660-5217, 4678-4738, 4678-4816, 4683-4816, 4687-4824, 4689-4998, 4689-5130, 4689-5287, 4801-5354, 4802-4985, 4812-5178, 4818-5380, 4889-5124, 4896-5198, 4904-5425, 4929-5381, 4981-5151, 4981-5249, 4981-5565, 4993-5228, 5005-5380, 5098-5521, 5126-5401, 5127-5434, 5270-5519, 5271-5560, 5324-5605, 5384-5802, 5386-5992, 5431-5703, 5431-5881, 5431-5939, 5431-5944, 5431-6023, 5469-5733, 5499-5939, 5521-5677, 5596-6077, 5630-5831, 5739-6288, 5745-5984, 5748-5999, 5779-6051, 5799-6068, 5860-6312, 5910-6472, 5940-6491, 5954-6435, 5954-6459, 6135-6600, 6139-6600, 6188-6520, 6193-6462, 6227-6470, 6289-6832, 6290-6487, 6319-6582, 6333-6584, 6348-6625, 6401-6679, 6401-6916, 6441-6698, 6441-6802, 6443-6705, 6449-6702, 6470-6727, 6480-6727, 6502-6792, 6534-6869, 6588-6876, 6591-6864, 6591-6875, 6628-6835, 6671-6929, 6707-6880, 6711-7010, 6751-7095, 6778-7063, 6843-7113, 6855-7102, 6880-7134, 6899-7153,
	6900-7186, 6947-7228, 6948-7200, 7030-7640, 7050-7638, 7087-7613, 7087-7622, 7091-7650, 7093-7634, 7105-7345, 7151-7631, 7161-7453, 7161-7649, 7161-7665, 7173-7650, 7176-7656, 7186-7659, 7192-7651, 7194-7650, 7199-7650, 7200-7653, 7202-7650, 7204-7650, 7214-7655, 7224-7650, 7228-7650, 7244-7498, 7249-7656, 7259-7650, 7268-7650, 7269-7655, 7269-7680, 7280-7650, 7292-7522, 7311-7650, 7316-7652, 7319-7650, 7322-7655, 7325-7650, 7331-7650, 7341-7650, 7353-7650, 7364-7659, 7380-7650, 7396-7653, 7399-7631, 7403-7647, 7412-7644, 7413-7663, 7414-7647, 7417-7607, 7417-7648, 7417-7661, 7417-7667, 7497-7650, 7502-7657, 7520-7652, 7533-7742, 7535-7656, 7592-7655
23/4259314CB1 / 1674	1-73, 1-121, 1-138, 1-217, 1-337, 1-341, 1-363, 1-418, 1-430, 1-440, 1-463, 1-470, 1-476, 1-484, 1-494, 1-508, 1-513, 1-521, 1-526, 1-628, 1-637, 2-252, 4-93, 4-246, 4-544, 4-588, 4-647, 4-681, 5-593, 5-604, 5-636, 16-526, 16-630, 20-476, 21-654, 77-680, 120-726, 198-778, 277-898, 411-1089, 442-1087, 534-1126, 535-1149, 551-1254, 562-819, 562-1060, 574-1166, 590-881, 591-1264, 670-1352, 673-1154, 677-1126, 696-1318, 721-1394, 726-980, 726-1151, 735-1336, 740-1456, 745-1126, 762-1456, 766-1252, 772-1213, 794-1456, 803-1289, 815-920, 887-1059, 893-1351, 897-1671, 900-1133, 927-1653, 961-1642, 968-1671, 1014-1633, 1033-1126, 1040-1641, 1054-1669, 1054-1674, 1055-1126, 1059-1126, 1059-1383, 1068-1650, 1161-1671, 1167-1671, 1179-1671, 1184-1674, 1191-1671, 1242-1656, 1259-1674

10

20

30

40

表4-2

ポリヌクレオチド SEQ ID NO/ Incyte ID/配列長	配列断片
24/3660046CB1 / 3671	1-827, 5-839, 11-599, 96-999, 237-558, 247-757, 348-932, 367-886, 376-2955, 438-700, 438-956, 633-1189, 648-1093, 828-1461, 955-1153, 1404-1591, 1660-2067, 1682-3032, 1709-2307, 1841-2388, 1925-2601, 1937-2463, 1989-2760, 2076-2486, 2167-2759, 2263-3046, 2441-2705, 2574-3040, 2634-2899, 2734-3297, 2736-3248, 2764-3070, 2843-3052, 2853-3220, 2860-3338, 2911-3511, 3007-3655, 3013-3461, 3086-3367, 3094-3325, 3114-3671
25/3016416CB1 / 2038	1-608, 170-2031, 371-608, 410-1096, 422-868, 429-963, 515-775, 515-1114, 519-1124, 579-926, 600-1266, 623-999, 623-1222, 661-1207, 661-1284, 731-1290, 762-1339, 852-961, 866-961, 947-1534, 963-1165, 1016-1544, 1026-1582, 1034-1339, 1068-1232, 1095-1383, 1273-1550, 1273-1786, 1301-1864, 1327-1540, 1407-1686, 1425-1880, 1445-2021, 1484-1991, 1534-2003, 1581-1894, 1635-1904, 1639-1928, 1682-1992, 1682-1999, 1682-2038, 1697-1960, 1697-1971
26/2133755CB1 / 2235	1-263, 69-584, 107-676, 107-718, 107-755, 109-300, 109-457, 109-480, 109-554, 109-566, 109-627, 109-657, 109-670, 109-720, 109-786, 109-788, 109-885, 111-200, 111-276, 111-377, 111-518, 111-560, 111-575, 111-577, 111-606, 111-610, 111-652, 111-661, 111-680, 111-697, 111-706, 111-749, 111-770, 111-782, 111-788, 111-840, 112-207, 114-697, 116-841, 117-260, 118-401, 118-576, 119-635, 127-543, 127-592, 128-808, 131-736, 132-371, 132-378, 132-444, 132-541, 132-601, 132-609, 209-400, 222-478, 283-542, 350-574, 365-779, 375-601, 405-608, 405-624, 633-890, 661-905, 697-964, 717-970, 720-939, 720-958, 720-968, 723-1039, 744-1011, 763-1055, 771-1001, 808-1276, 840-1264, 845-1161, 848-1085, 865-1167, 875-1204, 904-1177, 904-1329, 914-1180, 918-1198, 923-1194, 923-1366, 927-1384, 930-1210, 943-1155, 948-1156, 950-1200, 950-1226, 951-1501, 952-1254, 959-1200, 968-1287, 971-1186, 971-1206, 974-1200, 984-1255, 984-1553, 994-1276, 997-1264, 1025-1294, 1030-1232, 1034-1261, 1036-1263, 1036-1316, 1042-1250, 1046-1319, 1047-1313, 1051-1282, 1056-1327, 1057-1276,

10

20

30

40

表 4-3

ポリヌクレオチド SEQ ID NO/ Incyte ID/配 列長	配列断片
	1060-1340, 1064-1317, 1077-1328, 1082-1335, 1085-1354, 1127-1372, 1128-1563, 1135-1398, 1145-1402, 1145-1413, 1156-1458, 1167-1461, 1168-1333, 1180-1389, 1189-1438, 1191-1478, 1194-1509, 1225-1455, 1225-1646, 1227-1487, 1245-1507, 1260-1508, 1261-1524, 1269-1515, 1271-1506, 1273-1411, 1273-1516, 1277-1537, 1302-1483, 1307-1883, 1337-1556, 1338-1538, 1338-1544, 1350-1608, 1350-1619, 1370-1635, 1377-1692, 1384-1641, 1390-1676, 1393-1646, 1394-1623, 1399-1646, 1400-1635, 1409-1678, 1418-1721, 1419-1698, 1430-1708, 1438-1674, 1439-1673, 1439-1660, 1443-1706, 1460-1725, 1460-1983, 1528-1718, 1557-2206, 1564-1815, 1565-1762, 1607-1811, 1611-1840, 1627-2201, 1633-1816, 1643-2214, 1654-1914, 1704-2220, 1803-2026, 1820-2048, 1820-2055, 1850-2120, 1850-2220, 1890-2108, 1893-2113, 1893-2133, 2003-2206, 2048-2219, 2053-2235
27/5259957CB1 / 1851	1-346, 1-461, 1-478, 1-526, 1-569, 1-570, 53-550, 248-817, 259-709, 292-913, 312-820, 317-688, 346-919, 360-1014, 361-926, 455-1063, 510-1126, 526-1093, 539-1080, 623-1201, 630-1263, 632-1215, 690-1182, 696-1164, 696-1319, 768-1303, 785-1219, 885-1468, 888-1245, 898-1354, 905-1083, 915-1171, 966-1442, 1074-1673, 1108-1701, 1210-1851, 1226-1531, 1232-1743, 1300-1830
28/55029783CB1 / 1466	1-614, 110-817, 112-683, 117-675, 125-633, 126-647, 141-701, 141-765, 141-847, 141-851, 149-535, 149-700, 155-752, 155-766, 155-783, 155-784, 164-827, 165-789, 168-835, 168-855, 168-862, 170-644, 173-194, 207-779, 215-748, 236-434, 272-743, 405-608, 429-1116, 454-710, 454-938, 467-1116, 471-718, 487-1277, 500-734, 503-1116, 519-1132, 560-745, 560-1369, 561-1305, 564-1132, 638-1369, 683-1342, 977-1211, 977-1466, 1066-1249, 1081-1298
29/8032202CB1 / 1049	1-312, 1-655, 3-834, 256-974, 343-593, 343-594, 412-1049, 439-731, 439-909, 539-963

10

20

30

40

表 4-4

ポリスクレオチ ド SEQ ID NO/ Incyte ID/配 列長	配列断片
30/6937367CB1 / 2520	1-474, 3-121, 3-712, 44-314, 91-198, 91-206, 94-440, 175-600, 207-413, 210-675, 376-1202, 456-1090, 474-958, 524-1054, 533-1119, 554-746, 596-1190, 613-1202, 646-1054, 726-1108, 750-871, 760-1281, 783-903, 808-1056, 849-1300, 878-1441, 919-1678, 985-1242, 1027-1388, 1128- 1413, 1144-1688, 1158-1279, 1243-1512, 1243-1576, 1252-1449, 1268- 1769, 1269-1581, 1331-1593, 1348-1418, 1351-1388, 1386-1421, 1408- 1995, 1419-1463, 1419-1579, 1419-1631, 1419-1689, 1419-1707, 1419- 1755, 1419-1789, 1419-1822, 1419-1866, 1419-1872, 1419-1971, 1419- 2020, 1419-2058, 1436-1598, 1446-1637, 1452-1913, 1474-1664, 1474- 1666, 1480-1943, 1482-1692, 1510-1959, 1543-1681, 1546-2072, 1585- 2284, 1604-2123, 1604-2252, 1610-1762, 1617-1887, 1658-1926, 1716- 2018, 1752-1951, 1752-2229, 1753-2027, 1770-2054, 1822-2234, 1835- 2109, 1855-1996, 1855-2336, 1890-2187, 1933-2137, 1934-2497, 1979- 2253, 2019-2520, 2073-2312, 2092-2296, 2120-2347
31/3876510CB1 / 4360	1-619, 1-634, 6-322, 12-786, 14-140, 92-787, 147-463, 147-718, 147- 865, 147-879, 147-917, 147-920, 147-927, 147-931, 150-752, 150-860, 150-884, 164-473, 195-1051, 315-429, 315-532, 325-1145, 454-1042, 454- 1045, 592-1493, 610-1485, 630-757, 700-1454, 701-913, 744-1032, 767- 1457, 770-1457, 804-1126, 923-1154, 923-1375, 942-1197, 957-1239, 960- 1241, 1114-1792, 1155-1179, 1164-1441, 1165-1804, 1171-1198, 1236- 1703, 1249-1271, 1254-1588, 1321-1580, 1369-1865, 1456-1818, 1457- 1751, 1466-1819, 1523-1734, 1727-1983, 1736-2047, 1813-2114, 1839- 2115, 1858-2443, 1862-2154, 1863-2371, 1863-2384, 1863-2438, 1863- 2498, 1886-2165, 1924-2426, 2023-2365, 2114-2394, 2114-2693, 2114- 2805, 2114-2897, 2133-2663, 2176-2726, 2192-2473, 2200-2402, 2201- 2438, 2235-2504, 2250-2832, 2352-2696, 2364-2598, 2364-2885, 2364- 2910, 2364-2911, 2364-2976, 2364-2988, 2364-3032, 2364-3068, 2364- 3069, 2364-3172, 2388-2660, 2389-2749, 2392-2749, 2397-3276, 2405- 2696, 2444-3041, 2456-3159, 2463-2699, 2482-3112, 2490-3076, 2492- 2732, 2516-3037, 2544-3174, 2557-3214,

10

20

30

40

【0490】

【表4-5】

表4-5

ポリヌクレオチド SEQ ID NO/ Incyte ID/配列長	配列断片	
	2560-3109, 2562-2833, 2570-3277, 2588-3281, 2603-2816, 2606-2886, 2637-3118, 2640-3150, 2655-3152, 2676-3120, 2679-3362, 2694-3217, 2724-3161, 2728-3165, 2739-3205, 2744-3390, 2746-3418, 2756-3401, 2762-3217, 2762-3391, 2766-3182, 2768-3391, 2782-3398, 2801-3284, 2810-3007, 2810-3311, 2819-3289, 2856-3425, 2917-3425, 2942-3494, 2992-3445, 3005-3449, 3020-3672, 3029-3852, 3032-3652, 3040-3545, 3051-3594, 3063-3568, 3065-3555, 3080-3683, 3099-3612, 3113-3792, 3124-3613, 3152-3363, 3156-3626, 3194-3529, 3198-3466, 3213-3237, 3228-3792, 3252-3559, 3257-3971, 3266-3566, 3279-3535, 3296-3990, 3322-3849, 3335-3829, 3338-3956, 3349-3605, 3349-3793, 3349-3910, 3387-3889, 3404-4215, 3441-4333, 3455-4149, 3465-4055, 3479-4071, 3541-4037, 3553-4209, 3566-4133, 3589-3831, 3594-4209, 3646-4353, 3648-4223, 3649-4146, 3661-4161, 3687-3904, 3692-3944, 3693-4360, 3700-4104, 3706-3960, 3706-4157, 3706-4216, 3706-4250, 3706-4339, 3706-4355, 3723-3947, 3723-3950, 3734-3967, 3744-4323, 3749-4190, 3779-4273, 3794-4088, 3815-4346, 3835-4338, 3863-4346, 3868-4183, 3881-4357, 3891-4333, 3901-3936, 3907-4346, 3916-4345, 3918-4303, 3918-4358,	
	3921-4360, 3924-4346, 3933-4178, 3938-4358, 3953-4349, 3965-4347, 3986-4343, 3988-4360, 3998-4345, 4024-4343, 4035-4239, 4086-4354, 4133-4344, 4153-4342	
32/4900076CB1/ 3500	1-239, 1-242, 1-286, 101-246, 102-359, 105-273, 105-347, 105-407, 108-457, 196-312, 235-708, 319-519, 456-968, 516-642, 643-954, 643-1023, 643-1119, 645-1281, 651-1281, 673-1281, 854-1325, 854-1338, 854-1341, 854-1386, 854-1508, 857-1323, 859-1555, 916-1534, 1035-1636, 1040-1618, 1059-1335, 1061-1276, 1065-1624, 1107-1519, 1114-1201, 1114-1237, 1114-1250, 1114-1392, 1114-1459, 1114-1463, 1114-1464, 1114-1477, 1114-1481, 1116-1759, 1123-1632, 1135-1235, 1135-1237, 1135-1241, 1135-1248, 1135-1272, 1135-1290, 1135-1300, 1135-1308, 1135-1328, 1135-1340, 1135-1342, 1135-1370, 1135-1380, 1135-1403, 1135-1405, 1135-1423, 1135-1435, 1135-1436, 1135-1456, 1135-1464, 1135-1508, 1135-1546, 1135-1548, 1135-1560, 1135-1627, 1135-1631, 1135-1632, 1135-1666, 1135-1715, 1147-1201, 1148-1405, 1149-1592, 1158-1256, 1158-1264, 1158-1289, 1158-1294, 1158-1329, 1158-1336, 1158-1542, 1158-1575, 1158-1582, 1158-1659, 1160-1332, 1182-1465, 1182-1632, 1187-1473, 1188-1477, 1188-1716, 1211-1464, 1211-1477, 1217-1287, 1217-1288, 1217-1316, 1217-1323, 1217-1424,	

10

20

30

40

表 4-6

ポリヌクレオチ ド SEQ ID NO/ Incyte ID/配 列長	配列断片
	1217-1474, 1217-1540, 1217-1558, 1217-1575, 1217-1703, 1217-1728, 1217-1730, 1218-1405, 1218-1459, 1218-1477, 1218-1703, 1218-1772, 1219-1359, 1219-1405, 1219-1409, 1219-1416, 1219-1425, 1219-1465, 1219-1558, 1219-1592, 1219-1676, 1219-1722, 1233-1575, 1233-1591, 1235-1575, 1239-1733, 1240-1827, 1245-1477, 1249-1538, 1251-1575, 1263-1716, 1265-1883, 1266-1812, 1271-1416, 1271-1676, 1273-1518, 1281-1632, 1297-1772, 1298-1880, 1302-1541, 1302-1575, 1302-1735, 1302-1764, 1302-1812, 1303-1642, 1303-1687, 1303-1695, 1303-1772, 1303-1800, 1317-1632, 1319-1632, 1321-1575, 1323-1827, 1333-1918, 1347-1810, 1349-1967, 1356-1592, 1362-1592, 1362-1632, 1364-1884, 1368-1716, 1377-1896, 1383-1632, 1385-1832, 1386-1632, 1386-1876, 1387-1487, 1387-1522, 1387-1528, 1387-1575, 1387-1577, 1387-1584, 1387-1593, 1387-1594, 1387-1738, 1387-1759, 1387-1760, 1387-1870, 1387-1876, 1387-1884, 1387-2015, 1391-1968, 1401-1716, 1403-1716, 1404-1911, 1410-1632, 1417-1708, 1421-1718, 1422-1716, 1431-1884, 1441-2024, 1446-1676, 1446-1982, 1449-1810, 1461-1980, 1469-1716, 1469-1948, 1470-1716, 1470-1760, 1490-1716, 1490-1810, 1491-1984, 1492-2062, 1494-1600, 1494-1659, 1494-1661, 1494-1668, 1494-1676, 1494-1810, 1494-1832, 1494-1928, 1494- 1968, 1494-2012, 1495-1832, 1501-1928, 1509-1982, 1515-1735, 1515-1760, 1515-1810, 1521-1668, 1523-2135, 1532-1760, 1532-2064, 1533-1884, 1534-1672, 1534-1695, 1534-1738, 1534-1750, 1534- 1761, 1534-1762, 1534-1885, 1534-1927, 1534-1939, 1534-1947, 1534-2024, 1534-2052, 1538-2096, 1545-2064, 1549-2024, 1553-1764, 1553-1810, 1554-1810, 1554-2014, 1559-2239, 1569-1884, 1571- 1884, 1573-1810, 1575-1982, 1575-2079, 1585-1874, 1590-1884, 1608-2062, 1608-2146, 1609-1885, 1616-1844, 1616-2131, 1617-1982, 1633-2155, 1637-1884, 1638-1884, 1638-2131, 1638-2146, 1639- 1762, 1639-1827, 1639-1829, 1639-1836, 1639-1846, 1639-2002, 1639-2012, 1639-2122, 1639-2128, 1639-2136, 1643-2146, 1651-2011, 1652-2162, 1653-1982, 1655-1982, 1657-1884, 1659-2062, 1659- 2136, 1677-2294, 1683-1947, 1683-1969, 1683-1982, 1690-2294, 1691-2136, 1708-1928, 1708-2146, 1713-2150, 1722-1940, 1722-1982, 1722-2012, 1722-2098, 1742-2294, 1743-2146, 1744-2150, 1746- 1982, 1782-2291, 1785-2012, 1785-2062, 1785-2136, 1785-2146, 1786-1924, 1786-1942, 1786-1947, 1786-1993, 1786-1997, 1786-2002, 1786-2013, 1786-2014, 1786-2136, 1786-2149,

10

20

30

40

表 4 - 7

ポリヌクレオチド SEQ ID NO/ Incyte ID/配 列長	配列断片
	1786-2154, 1788-2012, 1788-2123, 1797-2136, 1801-2123, 1805-1863, 1805-1869, 1805-1911, 1805-1912, 1805-1920, 1805-1928, 1805-2014, 1805-2062, 1805-2095, 1805-2107, 1805-2150, 1805-2162, 1806-2062, 1806-2150, 1818-2294, 1823-2294, 1824-2136, 1825-2062, 1827-2136, 1837-2126, 1839-2294, 1851-2150, 1853-2162, 1860-2002, 1861-2136, 1866-2096, 1866-2136, 1869-1948, 1869-2146, 1888-2136, 1889-2294, 1890-2136, 1890-2137, 1890-2146, 1890-2162, 1890-2353, 1890-2433, 1891-1948, 1891-2020, 1891-2024, 1891-2079, 1891-2081, 1891-2088, 1891-2097, 1891-2131, 1891-2135, 1891-2136, 1891-2160, 1905-2146, 1907-2146, 1909-2294, 1913-2150, 1914-2136, 1914-2146, 1920-2150, 1929-2136, 1935-2136, 1935-2146, 1952-2162, 1956-2136, 1957-2149, 1957-2150, 1961-2149, 1965-2136, 1974-2088, 1974-2146, 1974-2160, 1974-2162, 1989-2150, 1991-2150, 1994-2136, 1995-2162, 1996-2136, 1998-2146, 2034-2136, 2037-2146, 2037-2150, 2037-2162, 2038-2125, 2040-2149, 2049-2136, 2053-2136, 2057-2150, 2058-2136, 2058-2150, 2076-2136, 2077-2150, 2079-2146, 2079-2149, 2094-2136, 2113-2149, 2256-2450, 2269-2529, 2269-2859, 2293-2585, 2372-2423, 2372-2429, 2389-2706,
	2391-2689, 2409-2668, 2463-2708, 2465-2715, 2518-2814, 2567-3163, 2727-3405, 2752-3404, 2766-3401, 2815-3067, 2818-3452, 2901-3133, 2901-3167, 2962-3457, 3003-3248, 3103-3487, 3276-3500
33/1543848CBI/1 366	1-352, 135-672, 146-819, 160-873, 174-428, 174-439, 174-440, 174-456, 185-304, 185-352, 185-359, 185-361, 185-366, 185-401, 185-423, 185-431, 185-442, 185-445, 185-451, 185-468, 185-521, 185-560, 186-298, 187-432, 190-435, 192-435, 194-414, 194-476, 199-303, 199-459, 200-430, 200-443, 200-453, 200-457, 200-465, 200-472, 201-373, 201-426, 203-359, 206-583, 207-499, 209-393, 212-491, 213-460, 213-465, 215-294, 215-352, 215-353, 221-442, 222-392, 222-518, 225-724, 227-583, 228-348, 230-490, 241-805, 243-491, 250-352, 251-442, 253-501, 253-513, 258-357, 258-522, 258-528, 258-534, 258-825, 259-357, 261-352, 261-500, 261-506, 261-554, 261-583, 262-873, 265-953, 269-546, 270-461, 274-877, 292-352, 292-534, 292-535, 292-557, 293-503, 293-583, 293-611, 301-588, 301-848, 306-464, 306-568, 307-583, 308-512, 312-768, 312-775, 312-1006, 350-432, 350-433, 350-439, 350-441, 350-512, 351-445, 351-472, 351-535, 351-548, 351-583, 351-687, 361-583, 365-997, 385-583, 414-917, 426-1072, 442-583, 448-583, 470-1154, 475-583, 511-1202, 518-583, 528-999, 530-1006, 543-583,

10

20

30

40

表 4-8

ポリスクレオチ ド SEQ ID NO/ Incyte ID/配 列長	配列断片
	552-583, 581-1261, 589-678, 593-843, 599-879, 609-671, 614-1164, 614-1175, 617-1177, 630-1177, 633-886, 633-888, 638-1154, 647-900, 650-1177, 671-911, 677-1361, 678-734, 680-734, 680-817, 680-819, 680-820, 680-835, 680-845, 680-849, 680-911, 680-913, 680-915, 680-931, 680-932, 680-1115, 680-1121, 683-1137, 684-1113, 688-1236, 697-1023, 697-1032, 705-964, 705-972, 707-971, 708-856, 718-982, 725-1330, 726-1137, 734-972, 735-1218, 735-1256, 735-1314, 737-1019, 752-1013, 758-1031, 758-1132, 758-1348, 758-1366, 780-1036, 791-1137, 797-1003, 797-1007, 797-1038, 797-1046, 797-1063, 797-1096, 799-1057, 799-1106, 819-1059, 822-1093, 834-1109
34/6254070CBI/ 4524	1-534, 1-680, 1-774, 1-4059, 286-1102, 418-1263, 427-525, 488-1069, 577-1268, 653-1269, 885-1536, 896-1149, 904-1193, 1000-1632, 1000-1656, 1006-1536, 1015-1289, 1069-1376, 1069-1536, 1279-1537, 1292-1514, 1576-1799, 1612-2128, 1664-2290, 1763-2014, 1805-2289, 1813-2397, 1931-2292, 1951-2020, 1956-2175, 1956-2505, 1956-2533, 1970-2236, 2119-2395, 2144-2437, 2220-2324, 2269-2978, 2342-2762, 2436-2996, 2488-3097, 2511-2951, 2538-2723, 2542-2816, 2544-2996, 2731-3347, 2755-3350, 2778-3288, 3070-3204, 3070-3377, 3070-3579, 3088-3359, 3096-3344, 3137-3291, 3188-3486, 3216-3652, 3259-3420, 3261-3539, 3281-3513, 3290-3663, 3327-3663, 3343-3363, 3364-3594, 3377-3590, 3469-3645, 3471-3663, 3503-3609, 3515-3663, 3536-3663, 3572-3663, 3584-3663, 3614-3663, 3615-3663, 3617-3663, 3653-3955, 3653-3981, 3653-4018, 3653-4097, 3658-3708, 3658-3816, 3658-3880, 3658-4077, 3662-3919, 3662-3938, 3674-4114, 3699-3783, 3699-3955, 3701-3807, 3746-4097, 3785-4097, 3800-4051, 3800-4097, 3803-4097, 3808-4159, 3833-4097, 3842-4097, 3901-4524, 3955-4097
35/1289839CBI/ 1157	1-699, 47-616, 198-624, 372-1157, 679-927, 918-1157
36/5565648CBI/ 1418	1-297, 1-510, 1-1380, 4-210, 7-693, 16-681, 24-251, 33-293, 41-519, 42-610, 42-674, 48-224, 59-257, 59-613, 71-350, 71-597, 72-399, 75-288, 77-385, 83-341, 84-373, 86-443, 86-492, 86-637, 91-384, 91-640, 91-660, 94-238, 94-345, 94-365, 94-580, 94-632, 96-517, 97-210, 97-395, 101-298, 101-299, 101-362, 102-341, 233-378, 270-540, 378-764, 450-741, 459-668, 463-662, 463-1028, 484-759, 518-764, 531-781, 532-825, 609-844, 634-891, 654-869, 665-890, 680-1127, 686-987, 686-992, 686-1028, 688-1323, 700-1322, 727-1375, 781-1418, 784-1398, 792-1338, 844-1329, 892-1333, 893-917, 893-973, 893-1111, 893-1191, 893-1339, 893-1418, 896-1329, 925-1204, 925-1219, 926-1172, 962-1226, 984-1233, 1127-1327, 1179-1336

10

20

30

40

表 4 - 9

ポリヌクレオチ ド SEQ ID NO/ Incyte ID/配 列長	配列断片
37/2764456CB1/ 4113	1-195, 37-597, 153-380, 168-597, 224-597, 239-834, 257-968, 305-546, 404-647, 539-759, 539-890, 541-1330, 610-834, 741-1269, 741-1365, 758-977, 1000-1470, 1000-1546, 1000-1674, 1138-1410, 1332-1956, 1343-1506, 1343-1508, 1405-1956, 1441-1773, 1459-1956, 1680-1929, 1686-2024, 1720-1988, 1722-2086, 1783-2262, 1818-2052, 2053-2632, 2059-2360, 2134-2591, 2134-2742, 2174-2472, 2174-2482, 2174-2607, 2175-2564, 2191-2822, 2196-2773, 2205-2500, 2238-2825, 2250-2690, 2324-2599, 2341-2772, 2402-3009, 2405-2471, 2430-2818, 2433-2998, 2442-2699, 2442-2893, 2453-2642, 2467-2516, 2505-2926, 2520-2765, 2520-2780, 2555-3147, 2560-3095, 2565-2840, 2569-3146, 2572-3146, 2588-2878, 2598-3194, 2676-3281, 2687-3164, 2717-3111, 2726-3376, 2780-3189, 2802-3111, 2859-3103, 2876-3569, 2890-3150, 2907-3454, 2920-3188, 2927-3171, 2959-3485, 3114-3687, 3116-3635, 3156-3535, 3156-3726, 3181-3478, 3191-3795, 3192-3487, 3195-3364, 3196-3449, 3199-3772, 3208-3795, 3214-3848, 3219-3721, 3223-3492, 3241-3475, 3241-3529, 3241-3814, 3242-3757, 3246-3476, 3250-3755, 3252-3536, 3252-3744, 3299-3550, 3299-3859, 3340-3877, 3342-3578, 3351-4033, 3357-3592, 3361-3751, 3380-4072, 3393-3688, 3401-3610, 3438-3915, 3453-4055, 3469-3704, 3472-4090, 3476-3847, 3487-4096, 3493-3708, 3515-4085, 3517-4094, 3524-4084, 3542-4064, 3545-3803, 3545-4103, 3551-3813, 3570-3814, 3570-4113, 3571-4092, 3585-3850, 3590-4102, 3593-3834, 3593-4063, 3601-4072, 3606-4110, 3620-3859, 3621-4112, 3623-3946, 3624-4101, 3626-3926, 3641-4108, 3648-4103, 3650-3890, 3656-4103, 3658-3947, 3660-4053, 3666-4113, 3669-4102, 3673-3849, 3678-3932, 3689-4103, 3693-4113, 3696-3920, 3705-4023, 3705-4103, 3711-3790, 3712-3762, 3712-3776, 3712-3778, 3712-3790, 3716-4104, 3719-3790, 3723-4104, 3727-3779, 3727-3788, 3732-4113, 3735-4107, 3739-4112, 3742-3790, 3748-3790, 3749-4102, 3787-4103, 3792-4105, 3803-3869, 3803-3881, 3810-3869, 3810-4021, 3818-3869, 3818-4100, 3823-3869, 3830-3879, 3833-4102, 3838-4113, 3850-4106, 3857-4113, 3874-4102, 3895-4103, 3977-4111

表 4-10

ポリヌクレオチド SEQ ID NO/ Incyte ID/配列長	配列断片
38/5734806CBI/ 7058	1-789, 1-7058, 454-485, 488-716, 488-959, 488-1017, 491-1035, 524-753, 713-1374, 1025-1321, 1236-1847, 1268-1516, 1284-1543, 1290-1695, 1302-1585, 1310-1860, 1318-2019, 1329-1557, 1350-1952, 1383-1561, 1390-1643, 1392-1943, 1403-1664, 1432-1804, 1483-1776, 1484-1789, 1484-1800, 1488-2101, 1489-1775, 1521-1777, 1562-1815, 1584-1881, 1597-1882, 1614-2220, 1636-2100, 1676-2074, 1697-1989, 1740-2046, 1741-2004, 1759-2059, 1768-2045, 1774-2079, 1832-2102, 1840-2386, 1845-2120, 1847-2088, 1853-2099, 1857-2138, 1888-2370, 2009-2277, 2040-2313, 2090-2586, 2092-2352, 2104-2386, 2108-2327, 2181-2434, 2200-2804, 2265-2962, 2270-2915, 2282-2926, 2294-2516, 2329-2595, 2332-2612, 2355-2582, 2359-2614, 2360-3027, 2445-2722, 2484-3041, 2525-3158, 2543-3133, 2636-2940, 2642-3238, 2647-3103, 2651-2932, 2678-2885, 2729-3077, 2737-3015, 2737-3117, 2739-3039, 2739-3103, 2740-3310, 2754-3368, 2807-3148, 2850-3308, 2876-3486, 2992-3498, 3079-3367, 3113-3665, 3119-3628, 3154-3806, 3183-3747, 3265-3880, 3272-3915, 3323-3949, 3410-3726, 3437-3898,
	3447-3874, 3448-3958, 3493-3781, 3493-3938, 3539-4077, 3597-4291, 3644-3937, 3694-3874, 3694-3993, 3750-3901, 3793-3972, 3970-4531, 4007-4268, 4073-4663, 4122-4365, 4139-4435, 4152-4607, 4260-4521, 4260-4754, 4368-4675, 4368-4829, 4400-4676, 4408-4628, 4411-4725, 4434-4875, 4522-4752, 4537-4772, 4539-4798, 4539-4994, 4565-4756, 4651-4889, 4653-4901, 4681-4935, 4717-4966, 4717-5203, 4717-5240, 4772-5022, 4802-5092, 4804-5073, 4813-5061, 4830-5126, 4831-5115, 4832-5084, 4908-5328, 4922-5225, 4925-5204, 4935-5234, 4950-5412, 4986-5328, 5069-5353, 5080-5313, 5096-5474, 5123-5381, 5161-5646, 5177-5328, 5204-5430, 5225-5431, 5243-5514, 5251-5488, 5251-5489, 5256-5481, 5292-5732, 5294-5543, 5321-5608, 5322-5608, 5325-5786, 5415-5790, 5450-5724, 5452-5806, 5483-5715, 5504-5780, 5544-5817, 5557-5824, 5584-5824, 5587-5825, 5599-5868, 5621-5825, 5646-5823, 5663-5823
39/7495168CBI/ 1380	1-423, 1-481, 359-697, 367-714, 367-1024, 369-707, 369-713, 369-714, 369-1000, 376-1038, 382-779, 384-755, 389-765, 395-624, 406-714, 407-697, 431-605, 431-664, 453-876, 458-951, 469-1082, 488-900, 491-1129, 495-949, 512-1107, 512-1182, 532-1183, 536-807, 540-1125, 559-1075, 559-1183, 561-1152, 575-804, 575-1152, 582-961, 590-1162, 636-1067, 640-907, 659-926, 661-1221, 668-1252, 701-927, 702-1027, 705-1239, 724-955, 728-1017, 796-1049, 888-1327, 891-1114, 927-1121, 931-1380

10

20

30

40

【 0 4 9 6 】

【 表 4 - 1 1 】

表 4-11

ポリヌクレオチド SEQ ID NO/ Incyte ID/配列長	配列断片
40/7483131CB1/1773	1-600, 4-320, 19-510, 45-464, 53-291, 60-642, 63-336, 63-586, 163-447, 189-447, 255-567, 294-812, 294-830, 403-623, 460-737, 473-951, 475-617, 516-642, 572-750, 606-854, 616-886, 628-876, 693-987, 695-972, 757-1354, 869-1101, 870-1400, 886-1131, 925-1675, 931-1166, 931-1441, 963-1559, 992-1308, 997-1269, 1020-1296, 1038-1357, 1098-1669, 1105-1711, 1172-1422, 1176-1701, 1185-1685, 1191-1332, 1192-1709, 1192-1713, 1195-1332, 1252-1723, 1255-1723, 1258-1721, 1274-1721, 1289-1728, 1290-1500, 1290-1645, 1295-1721, 1300-1705, 1302-1773, 1306-1585, 1306-1724, 1307-1584, 1311-1447, 1311-1734, 1316-1722, 1332-1724, 1336-1724, 1338-1721, 1339-1574, 1347-1610, 1347-1721, 1347-1724, 1349-1590, 1393-1717
41/4558650CB1/2810	1-444, 78-452, 79-509, 100-453, 140-799, 163-462, 359-823, 455-699, 455-722, 455-818, 455-869, 455-887, 455-931, 455-938, 455-944, 456-874, 460-1062, 464-879, 464-925, 464-985, 464-989, 464-1003, 464-1058, 466-905, 466-1129, 469-933, 473-942, 482-940, 489-880, 491-647, 491-673, 491-798, 499-877, 528-710, 642-1253, 643-1159, 842-1158, 900-1320, 907-1181, 929-1297, 930-1180, 930-1316, 955-1315, 958-1313, 983-1325, 993-1301, 998-1256, 999-1611, 1030-1290, 1064-1309, 1078-1280, 1078-1325, 1086-1283, 1107-1369, 1125-1309, 1125-1325, 1141-1325, 1332-1931, 1333-1731, 1334-1589, 1355-1624, 1361-1593, 1376-1721, 1385-1730, 1394-1797, 1394-1943, 1411-1876, 1438-1705, 1438-2022, 1439-1685, 1471-2032, 1498-1616, 1560-1834, 1600-1985, 1612-2088, 1622-2278, 1675-1928, 1779-2342, 1851-2345, 1855-2128, 1856-2119, 1915-2565, 1926-2532, 1933-2230, 1933-2261, 1942-2531, 1945-2453, 1945-2482, 1961-2531, 1965-2174, 1989-2766, 1996-2602, 1997-2488, 1999-2488, 2011-2459, 2041-2242, 2071-2300, 2086-2351, 2113-2373, 2114-2348, 2116-2368, 2116-2633, 2117-2715, 2133-2325, 2140-2363, 2146-2665, 2147-2766, 2151-2410, 2184-2731, 2259-2778, 2286-2602, 2324-2810, 2331-2465, 2344-2596, 2346-2810, 2356-2810, 2401-2474

表 4 - 1 2

ポリヌクレオチ ド SEQ ID NO/ Incyte ID/配 列長	配列断片
42/7506195CBI/ 2549	1-234, 11-454, 11-2549, 88-184, 88-310, 88-459, 88-462, 88-466, 89-519, 110-199, 110-463, 150-809, 164-462, 173-472, 279-462, 369-833, 465-709, 465-732, 465-828, 465-879, 465-897, 465-941, 465-948, 465-954, 466-884, 470-1072, 474-889, 474-935, 474-995, 474-999, 474-1013, 474-1068, 476-915, 476-1139, 479-943, 483-952, 492-950, 499-890, 501-657, 501-683, 501-808, 509-887, 538-720, 653-1159, 760-1358, 852-1157, 949-1325, 1155-1503, 1158-1366, 1158-1397, 1158-1494, 1167-1570, 1167-1716, 1184-1649, 1211-1478, 1211-1795, 1212-1458, 1244-1805, 1271-1389, 1314-1905, 1315-1722, 1315-1877, 1316-1927, 1333-1607, 1373-1758, 1381-1726, 1381-1737, 1385-1861, 1448-1701, 1508-2074, 1552-2115, 1605-1777, 1624-2118, 1628-1901, 1629-1892, 1699-2305, 1706-2003, 1706-2034, 1715-2304, 1718-2226, 1718-2255, 1734-2304, 1738-1947, 1742-2099, 1751-2253, 1762-2539, 1769-2221, 1769-2375, 1770-2261, 1772-2261, 1784-2232, 1814-2015, 1844-2073, 1859-2124, 1886-2146, 1887-2121, 1889-2141, 1889-2406, 1890-2488, 1906-2098, 1913-2136, 1920-2539, 1924-2183, 1944-2236, 1957-2504, 2032-2549, 2059-2375, 2069-2238, 2097-2549, 2104-2238, 2117-2369, 2174-2247

表 5

ポリスクレオチ ド Seq ID No:	Incyte プロジ エクト ID:	代表的ライブラリ
22	1351608CB1	PGANNT01
23	4259314CB1	STOMTMR02
24	3660046CB1	BRAIUNF01
25	3016416CB1	THYRDIE01
26	2133755CB1	SINTFER02
27	5259957CB1	KIDETXS02
28	55029783CB1	BLADTUT04
29	8032202CB1	TESTNOT11
30	6937367CB1	FTUSTUR01
31	3876510CB1	HEARFET02
32	4900076CB1	HNT2TXT01
33	1543848CB1	MPHGNOT03
34	6254070CB1	LUNPTUT02
35	1289839CB1	BRAHTDR04
36	5565648CB1	LIVRFET05
37	2764456CB1	COLNNT03
38	5734806CB1	THYMNOR02
39	7495168CB1	NOSETUR01
40	7483131CB1	KIDNNOT19
41	458650CB1	BRAUNOR01
42	7506195CB1	CONNTUT05

【 0 4 9 9 】
【 表 6 - 1 】

表6-1

ライブラリ	ベクター	ライブラリの説明
BLADTUT04	pINCY	ライブラリは、根治的膀胱切除、前立腺除去および輸精管切除時に60才白人男性から取り除かれた膀胱腫瘍組織から単離されたRNAを用いて作製された。病理は左膀胱壁にグレード3の移行上皮癌が見られた。ドームと三角部に原位癌が同定された。患者の病歴には喫煙がある。家族歴にはI型糖尿病、胃の悪性新生物、アテローム性冠動脈疾患および急性心筋梗塞がある。
BRAHTDR04	PCDNA2.1	このランダムプライムされたライブラリは、胆管癌で死亡した55才白人女性から切除した旧皮質、前海馬組織から単離したRNAを用いて作製した。病理は、門蓋部全体に優勢な軽度の髄膜線維形成があり、帯状皮質白質および視床に軸索球状体が散在し、内嗅皮質および中脳水道灰白質領域に、少数の神経原線維変化が散在した。随伴する腫瘍組織の病理は、残存腫瘍または再発性腫瘍を持つ高分化型の肝臓胆管癌を示した。患者の病歴には、胆管癌、術後パッド・キアーリ症候群、胆汁性腹水、水胸症、脱水症、栄養不良、乏尿、および急性腎不全がある。過去の手術には、胆嚢切除術および、肝臓の85%の摘除がある。
BRAIUNF01	PRARE	この5' キャップの単離した完全長ライブラリは、69才の白人男子から除去した脳腫瘍に由来のDU145細胞株から単離したRNAを用いて作製した。細胞は、14時間、無処理であった。病理は転移性病を示した。患者の病歴は、リンパ球性白血病(3年間)と、前立腺癌(脳に転移)がある。
BRAUNOR01	pINCY	このランダムプライムされたライブラリは、アテローム性動脈硬化による出血および胸部大動脈破裂で死亡した81才白人女性から得た、線条体、淡蒼球、および後部皮層組織から単離したRNAを用いて作製した。病理は、中等度のアテローム性動脈硬化(内頸動脈を同側性に侵す)、微限的梗塞(前頭皮質および海馬)、並びに年齢に一致した散在性びまん性のアミロイド病および神経原線維変化を示した。肉眼的には、軟膜は上矢状静脈洞に沿った軽度の肥厚と硝子質化とを示すのみであった。残りの軟膜は薄かった。また、充血した血管がいくらか見られた。主に前頭葉および前頭葉、及び側頭葉(両側)に軽度の萎縮が見られた。微視的には、アルツハイマー2型星状細胞のベア複数の新皮質の深層にあった。中前頭皮質の深部灰白質のニューロンの周りでサテリトース(satellitosis)が増大していた。扁桃体には微量のびまん性斑と神経原線維変化とがあった。

【0500】
【表6-2】

表 6－2

ライブラリ	ベクター	ライブラリの説明
		後部海馬には、反応性グリオシスに囲まれたヘモジデリン含有マクロファージを伴う、微細的な囊胞性の空洞化領域を含んでいた。患者の病歴は、敗血症、胆管炎、術後無気肺、肺炎冠動脈疾患 (CAD)、左室肥大に起因する心肥大、巨脾症、細動脈性腎硬化、結節性コロイド甲状腺腫、気腫、鬱血心不全 (CHF)、甲状腺機能低下症および末梢血管疾患を含んでいた。
COLNNOT23	pINCY	ライブラリは 16 才の白人男性の全結腸切除 (腹会陰式切開を併行) 時に採取した罹患結腸組織から単離した RNA を用いて作製した。病理は胃炎と全結腸炎 (pancolitis、急性期の潰瘍性大腸炎に一致) を示した。炎症は横行結腸でより重度であり、炎症は粘膜に限られた。上行結腸と S 状結腸の併発は軽度であり、盲腸、直腸、回腸終端部の顕著な併発はなかった。家族歴には過敏性腸症候群がある。
CONNTUT05	pINCY	ライブラリは、頭蓋病変切除時に 34 才白人女性から取り除かれた腫瘍性の頭蓋軟組織から単離された RNA を用いて作製された。病理は、真皮と皮下組織へのインプラントを形成するグレード 3 上衣細胞腫を示し、高密度の線維形成を伴っていた。患者の病歴には発作、骨痛、脳癌がある。手術には頭蓋形成と脳腫瘍切除があり、治療には全脳放射線 (whole brain radiation) があった。家族歴には不安と抑鬱がある。
FTUBTUT01	PCDNA2.1	このランダムプライムのライブラリは 85 歳の白人女性の両側性卵管卵巣摘出および子宮摘出中に採取した卵管腫瘍組織から単離された RNA を使って作製された。病理は、低分化型の類内膜型 (80%) および漿液性 (20%) 混合腺癌を示し、癌は粘膜に留まり、子宮壁の併発はなかった。 in situ の類内膜腺癌もあった。関連する子宮腫瘍の病理学検査によると、子宮内膜ポリープ内に in situ の限局性類内膜腺癌および浸潤性の中分化腺癌が生じていた。子宮腫瘍に、転移性の類内膜型および漿液性の腺癌があった。患者の病歴は、甲状腺の髄様癌と心筋梗塞があった。
HEARFET02	pINCY	ライブラリは妊娠 23 週目で死産の左心低形成症胎児 (白人男子) から採取した心組織から単離した RNA を用いて作製した。

【 0 5 0 1 】
【 表 6－3 】

表 6－3

ライブラリ	ベクター	ライブラリの説明
HNT2TXT01	pINCY	ライブラリは、hNT2 細胞株 (分化決定済み (committed) 神経前駆体の特性を示したヒト奇形癌に由来) から単離した RNA を使用して作製された。細胞の処理を、マウスのレプチン (1 μ /ml) と 9 cis レチノイン酸 (3.3 μ M) とで 6 日間おこなった。
KIDETXS02	pINCY	このサブトラクショナル形質転換胚細胞株ライブラリは、腎臓上皮組織から得られた処理済の形質転換胚細胞株 (293-EBNA) に由来する 900 万のクローンと、腎臓上皮組織ライブラリ由来の無処理の形質転換胚細胞株 (293-EBNA) からの 190 万のクローンとの 2 ラウンドのサブトラクショナルハイブリダイゼーションにかけて作製された。サブトラクショナルの開始ライブラリは処理済の形質転換された胚細胞株 (293-EBNA) から単離された RNA を使って作製された。これらの細胞は 5 アザ 2'-デオキシシチジンで処理され、アデノウイルス 5 の DNA で形質転換された。サブトラクショナルのためのハイブリダイゼーションプロローブは同じ細胞株からの無処理の 293-EBNA 細胞から単離された RNA から同様に作製されたライブラリから得られた。サブトラクショナルハイブリダイゼーション条件は Swaroop 他、NAR 19 (1991):1954 および Bonaldo 他 Genome Research (1996) 6:791 の方法論に基づく。
KIDNNOT19	pINCY	ライブラリは 65 才の白人男子の試験開腹および腎尿管膀胱切除時に採取した腎臓組織から単離した RNA を用いて作製した。随伴する腫瘍組織の病理学検査では、グレード 1 の腎細胞癌が左腎の上極内に見られた。患者の病歴には腹部皮膚の悪性黒色腫、結腸の良性腫瘍、脳血管障害および臍ヘルニアがある。家族歴には心筋梗塞、アテローム性冠動脈疾患、脳血管障害、前立腺癌、心筋梗塞およびアテローム性冠動脈疾患がある。
LIVRFET05	pINCY	ライブラリは妊娠 20 週で死亡の白人男子胎児の肝臓から採取した CD34+前駆細胞組織から単離した RNA を用いて作製された。

【 0 5 0 2 】
【 表 6－4 】

表 6 - 4

ライブラリ	ベクター	ライブラリの説明
LUNPTUT02	pINCY	ライブラリは、55歳の白人女性の肺全摘除中に採取した肋膜腫瘍組織から単離したRNAを使用して作製された。病理学的検査では平滑筋肉腫と最もよく一致するグレード3の肉腫があり、子宮がライマリーで、右下葉および中葉の一部を置き換える <i>bosellated</i> 塊を形成していた。腫瘍は隣接する壁側胸膜および心臓に及んでいた。腫瘍を構成する多数の小結節はほとんど完全な壊死を示していた。右上葉は無気性だが腫瘍の併発は無かった。細胞小結節のマイクロセクションは、活発な有糸分裂活性を示していた。心臓は直接関与していたが、そのマージンには腫瘍がなかった。平滑筋アクチンは陽性だった。エストロゲン受容体は陰性で、プロゲステロン受容体は陽性だった。患者の病歴としては、息切れ、消化性潰瘍、肺病、子宮癌、正常な出産、タバコの濫用および欠乏性貧血があった。以前の手術としては、肺の病巣の内視鏡切除があった。家族の病歴としては、アテローム硬化型冠動脈疾患、乳癌、タイプII糖尿病、および多発性硬化症等があった。
MPHGN03	PBLUESCRIPT	ライブラリは互いに関係のない男性と女性のドナーから得られたバフィコートユニットから単離された増殖性接着性単核細胞から単離されたRNAを使って構築された。
NOSETUE01	PCDNA2.1	このライブラリはランダムプライムライブラリは、全額骨切除（再建、鼻切除、および開頭を併行）時に45歳の白人男性から除かれた鼻部および篩状腫瘍 (cribriform tumor) 組織から単離したRNAを用いて作製された。病理学検査では、鼻腔と篩状領域に嗅覚神経芽腫 (olfactory neuroblastoma) が見られた。患者には頭、顔、首の癌と鼻出血が見られた。患者の病歴としては、外因性喘息、頭、顔、首の癌および鼻出血があった。過去の手術としては再建を伴う全額骨切除が含まれる。患者の使用薬剤としては <i>Biaxin</i> , <i>Atessalon</i> , および <i>Vallium</i> 等があった。患者は放射線療法を受けた。家族歴としては、母親に慢性的な副鼻腔炎、父親にタイプIIの糖尿病があった。
PGANN01	PSPORT1	ライブラリは、46歳の白人男性の試験開腹中に腹内領域から採取したパラガングリオン性腫瘍組織から単離したRNAを用いて作製された。病理学検査では、良性の傍神経節腫を示し、グレード2の腎細胞癌 (明細胞タイプ、被膜に貫入していない) に付随していた。外科的縁 (surgical margin) では、腫瘍が陰性であった。
SINTPER02	pINCY	このランダムプライムライブラリは、死亡した白人男子胎児から採取した小腸組織から単離したRNAを用いて作製された。死因は胎児性死亡である。

10

20

30

【 0 5 0 3 】

【 表 6 - 5 】

表 6-5

ライブラリ	ベクター	ライブラリの説明
STOMTMR02	PCDNA2.1	このランダムプライムされたライブラリは 76 歳白人男性の噴門側胃切除および部分食道切除時に摘出された罹患胃組織から単離した RNA を用いて作製した。病理検査は、慢性胃炎を示した。マッチした腫瘍組織の病理は浸潤性でグレード 3 の腺癌は胃食道接合部に、潰瘍化した塊を形成していた。腫瘍の浸潤は固有筋層を通過して食道周囲脂肪組織にまで及んでいた。胃周囲の 4 の内 1 つのリンパ節が腫瘍に陽性であった。患者の病歴には嚥下障害、アテローム性冠動脈疾患、皮膚の悪性メラノーマ、COPD、良性の大腸腫瘍、肺上葉の悪性新生物、およびアルコール濫用が含まれる。家族歴には、アテローム性冠動脈疾患および心筋梗塞を含む。
TESTNOT11	pINCY	ライブラリは総死した 16 歳の白人男子から採取した精巣組織から単離した RNA を用いて作製した。患者の病歴は薬物使用 (タバコ、マリファナおよびコカイン) を含み、使用薬剤としてはリチウム、リタリンおよび Paxil 等があった。
THYMNOR02	pINCY	ライブラリは、2 歳白人女子の胸腺摘出および左心房室フィステルのパッチ閉鎖中に除去した胸腺組織から単離した RNA を用いて作製された。病理学検査では、胸腺に肉眼的異常は見られなかった。患者には先天性心異常が見られた。患者の病歴には、左室性単心室 (double inlet left ventricle)、痕跡右心室、肺高血圧症、チアノーゼ、大動脈弁下部狭窄、発作、頭蓋底の骨折がある。家族歴には逆流性腎症 (reflux neuropathy) がある。
THYRDIE01	PCDNA2.1	この 5' に偏向してランダムプライムされたライブラリは、22 歳の白人女性の甲状腺の閉鎖式生検、部分的甲状腺切除、および所属リンパ節切除時に摘出した罹患甲状腺組織から単離した RNA を用いて作製した。病理学検査では腺腫性過形成が見られた。患者は甲状腺の悪性腫瘍を示した。患者の病歴は、正常分娩とアルコール濫用、タバコ濫用があった。以前に鼓膜切開術を受けている。患者の使用薬剤には、非特定タイプ II の終口避妊薬が含まれる。家族歴には、母の高脂血症と鬱病性障害、祖父母の良性高血圧症、鬱病性心不全、慢性白血球病が含まれる。

【 0 5 0 4 】
【 表 7 - 1 】

【 0 5 0 5 】
【 表 7 - 2 】

表 7 - 1

プログラム	説明	参考文献	パラメータ閾値
ABI FACTURA	核酸配列においてベクター配列を除去して、あいまいな塩基をマスクするプログラム。	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	アミノ酸配列または核酸配列の比較、注釈に有用なファーストデータファインダ (Fast Data Finder)。	Applied Biosystems, Foster City, CA. Paracel Inc., Pasadena, CA.	不一致<50%
ABI AutoAssembler	核酸配列を構築するプログラム。	Applied Biosystems, Foster City, CA.	
BLAST	アミノ酸配列および核酸配列の配列類似性検索に有用なベージックローカルアライメント検索ツール (Basic Local Alignment Search Tool)。BLASTには5つの機能がある: blastp, blastn, blastx, tblastnおよびtblastx。	Altschul, S.F. 他 (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. 他 (1997) (1997) Nucleic Acids Res. 25:3389-3402.	EST: 確率値=1.0E-8以下 完全長配列: 確率値= 1.0E-10以下
FASTA	一群の同じタイプの配列と問合せ配列との間の類似性について検索するPearson およびLipman アルゴリズム。FASTAは最少5つの機能からなる: fasta, ifasta, fastx, tfastxおよびksearch。	Pearson, W.R. 及びD.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448, Pearson, W.R. (1990) Methods Enzymol. 183:63-98; およびSmith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	EST: fasta E 値=1.06E-6 構築された ESTs: fasta 同一性= 95%以上 一致した長さ=200塩基以上 fastx E 値=1.0E-8 以下 完全長配列: fastx スコア=100以上
BLIMPS	配列をBLOCKS、PRINTS、DOMO、PRODOM およびPFAM データベースの配列と対応させて遺伝子ファミリー、配列相同性および構造的フィンガープリント領域を検索するBLoCKs IMProved Searcher。	Henikoff, S. 及びJ. G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572, Henikoff, J. G. 及びS. Henikoff (1996) Methods Enzymol. 266:88-105, Attwood, T. K. 他. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	確率値= 1.0E-3以下
HMMER	問合せ配列をPFAMのようなタンパク質ファミリーコンセンサス配列の隠れマルコフモデル (HMM) に基づいたデータベースに対して検索するアルゴリズム。	Krogh, A. 他. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. 他 (1998) Nucleic Acids Res. 26:320-322; Durbin, R. 他. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, 1-350 ページ。	PFAM ヒット: 確率値= 1.0E-3 以下 シグナル ペプチド ヒット: スコア=0以上

【 0 5 0 6 】
【 表 8 】

表 7 - 2

プログラム	説明	参考文献	パラメータ閾値
ProfileScan	Prositeで定義された配列パターンと一致するタンパク質配列内の構造的モチーフおよび配列モチーフを検索するアルゴリズム。	Gribskov, M. 他 (1988) CABIOS 4: 61-66, Gribskov, M. 他 (1989) Methods Enzymol. 183: 146-159, Bairoch, A. 他. (1997) Nucleic Acids Res. 25:217-221.	ノーマライズされた質スコアと特定のPrositeモチーフに対する GCG指定「HIGH」値。 通常、スコア=1.4~2.1.
Phred	高い感度と確率で自動配列決定機トレースを調べるベースコーリングアルゴリズム。	Ewing, B. 他 (1998) Genome Res. 8:175-185, Ewing, B及びP. Green (1998) Genome Res. 8:186-194.	
Phrap	Smith-Waterman アルゴリズムの効率的なインプリメンテーションに基づくSWATやCrossMatchを含むPhils Revised Assembly プログラムで、配列相同性の検索やDNA配列の構築に有用である。	Smith, T.F. 及び M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. 及び M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	スコア=120以上 一致した長さ=56以上
Consed	Phrap 構築物の表示および編集用グラフィカルツール。	Gordon, D. 他 (1998) Genome Res. 8:195-202.	
SPScan	タンパク質配列をスキャンして、分泌シグナルペプチドの存在を調べる重み行列解析プログラム。	Nielson, H. 他 (1997) Protein Engineering 10:1-6; Claverie, J.M. 及び S. Audic (1997) CABIOS 12:431-439.	スコア=3.5以上
TMAP	重み行列を用いて蛋白配列での膜貫通セグメントを描写し配向を決定するプログラム。	Persson, B. およびP. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. および P. Argos (1996) Protein Sci. 5:363-371.	
TMHMM	隠れMarkovモデル(HMM) を使ってタンパク質配列上の膜貫通セグメントを描写し、配向を決定するプログラム。	Sonnhammer, E.L. 他 (1998) Proc.Sixth Intl. Conf. on Intelligent Systems for Mol., Biol., Glasgow 他, 編集, The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, 175-182ページ.	
Motifs	Prositeで定義されたものと一致するパターンのアミノ酸配列を検索するプログラム。	Bairoch, A. 他 (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, バージョン9, M51-59ページ, Genetics Computer Group, Madison, WI	

表 8

SEQ ID NO.	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST アレル	アレル 1	アレル 2	アミノ酸	白人アレル 1 頻度	アフリカ系アレル 1 頻度	アジア系アレル 1 頻度	ヒスパニック系アレル 1 頻度
42	7506195	6753167J1	SNP001210 36	111	584	A	A	G	Q145	n/d	n/a	n/a	n/a
42	7506195	6763024J1	SNP001210 36	111	584	A	A	G	Q145	n/d	n/a	n/a	n/a
42	7506195	6767449J1	SNP001210 36	110	584	G	A	G	R145	n/d	n/a	n/a	n/a
42	7506195	6769029J1	SNP001210 36	109	584	G	A	G	R145	n/d	n/a	n/a	n/a
42	7506195	7754321H1	SNP001210 36	215	420	A	A	G	P90	n/d	n/a	n/a	n/a

【国際公開パンフレット】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number
WO 02/097032 A2

(51) International Patent Classification: C12N

(21) International Application Number: PCT/US02/11152

(22) International Filing Date: 5 April 2002 (05.04.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/282,110	6 April 2001 (06.04.2001)	US
60/283,294	11 April 2001 (11.04.2001)	US
60/286,820	26 April 2001 (26.04.2001)	US
60/287,228	27 April 2001 (27.04.2001)	US
60/291,662	16 May 2001 (16.05.2001)	US
60/291,846	18 May 2001 (18.05.2001)	US
60/293,727	25 May 2001 (25.05.2001)	US
60/295,340	1 June 2001 (01.06.2001)	US
60/295,263	1 June 2001 (01.06.2001)	US
60/349,705	15 January 2002 (15.01.2002)	US

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): AZIMZAI, Valda [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). AU-YOUNG, Janice, K. [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). BATRA, Sajeev [US/US]; 555 El Camino Real #709, Sunnyvale, CA 94087 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). BECHA, Shanya, D. [US/US]; 21062 Gary Drive #117, Castro Valley, CA 94546 (US). BOROWSKY, Mark, L. [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). BURFORD, Neil [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). DING, Li [CN/US]; 3353 Alma Street #146, Palo Alto, CA 94306 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Pollon Place Way, San Jose, CA 95121 (US). EMERLING, Brooke, M. [US/US]; 1735 Woodland Avenue #71, Palo Alto, CA 94303 (US). GANDHI, Ameena, R. [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). GIETZEN, Kimberly, J.

[US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). GRIFFIN, Jennifer, A. [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). HAFALIA, April, J.A. [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). HONCHELL, Cynthia, D. [US/US]; 400 Laurel Street #203, San Carlos, CA 94070 (US). LAL, Preet, G. [US/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). LEE, Soo, Yoon [KR/US]; 40 Westdale Avenue, Daly City, CA 94015 (US). LU, Dyung, Aina, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). ARVIZU, Chandra, S. [US/US]; 1706 Morocco Drive, San Jose, CA 95125 (US). RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). REDDY, Roopa [IN/US]; 1233 West McKinley Avenue #3, Sunnyvale, CA 94086 (US). SANJANWALA, Madhu, M. [US/US]; 210 Silvia Court, Los Altos, CA 94024 (US). TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). WALIA, Narinder, K. [US/US]; 890 Davis Street A #205, San Leandro, CA 94577 (US). WANG, Yumei, E. [CN/US]; 4624 Strawberry Park Drive, San Jose, CA 95129 (US). WARREN, Bridget, A. [US/US]; 2250 Homestead Court #2, Los Altos, CA 94024 (US). XU, Yuming [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). YANG, Junming [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). YAO, Montique, G. [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). YUE, Henry [US/US]; 826 Luis Avenue, Sunnyvale, CA 94087 (US). ZEBARJADIAN, Yeganeh [IR/US]; 830 Junipero Serra Boulevard, San Francisco, CA 94127 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GI, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,

[Continued on next page]

(54) Title: PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH

(57) Abstract: The invention provides human proteins associated with cell growth, differentiation, and death (CGDD) and polynucleotides which identify and encode CGDD. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of CGDD.



WO 02/097032 A2

WO 02/097032 A2 

GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BI, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NI, SN, TD, TG). *For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

Published:
— *without international search report and to be republished upon receipt of that report*

WO 02/097032

PCT/US02/11152

PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH**TECHNICAL FIELD**

This invention relates to nucleic acid and amino acid sequences of proteins associated with cell growth, differentiation, and death and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative disorders including cancer, developmental disorders, neurological disorders, autoimmune/inflammatory disorders, reproductive disorders, and disorders of the placenta, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of proteins associated with cell growth, differentiation, and death.

10

BACKGROUND OF THE INVENTION

Human growth and development requires the spatial and temporal regulation of cell differentiation, cell proliferation, and apoptosis. These processes coordinately control reproduction, aging, embryogenesis, morphogenesis, organogenesis, and tissue repair and maintenance. At the cellular level, growth and development is governed by the cell's decision to enter into or exit from the cell division cycle and by the cell's commitment to a terminally differentiated state. These decisions are made by the cell in response to extracellular signals and other environmental cues it receives. The following discussion focuses on the molecular mechanisms of cell division, embryogenesis, cell differentiation and proliferation, and apoptosis, as well as disease states such as cancer which can result from disruption of these mechanisms.

20

Cell Cycle

Cell division is the fundamental process by which all living things grow and reproduce. In unicellular organisms such as yeast and bacteria, each cell division doubles the number of organisms. In multicellular species many rounds of cell division are required to replace cells lost by wear or by programmed cell death, and for cell differentiation to produce a new tissue or organ. Progression through the cell cycle is governed by the intricate interactions of protein complexes. This regulation depends upon the appropriate expression of proteins which control cell cycle progression in response to extracellular signals, such as growth factors and other mitogens, and intracellular cues, such as DNA damage or nutrient starvation. Molecules which directly or indirectly modulate cell cycle progression fall into several categories, including cyclins, cyclin-dependent protein kinases, growth factors and their receptors, second messenger and signal transduction proteins, oncogene products, and tumor-suppressor proteins.

25

30

Progression through the cell cycle is governed by the intricate interactions of protein

WO 02/097032

PCT/US02/11152

complexes. This regulation depends upon the appropriate expression of proteins which control cell cycle progression in response to extracellular signals, such as growth factors and other mitogens, and intracellular cues, such as DNA damage or nutrient starvation. Molecules which directly or indirectly modulate cell cycle progression fall into several categories, including cyclins, cyclin-dependent protein
 5 kinases, growth factors and their receptors, second messenger and signal transduction proteins, oncogene products, and tumor-suppressor proteins.

The entry and exit of a cell from mitosis is regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins act by binding to and activating a group of cyclin-dependent protein kinases (Cdks) which then phosphorylate and activate selected proteins involved in
 10 the mitotic process. Cyclins are characterized by a large region of shared homology that is approximately 180 amino acids in length and referred to as the "cyclin box" (Chapman, D.L. and Wolgemuth, D.J. (1993) Development 118:229-40). In addition, cyclins contain a conserved 9 amino acid sequence in the N-terminal region of the molecule called the "destruction box". This sequence is believed to be a recognition code that triggers ubiquitin-mediated degradation of cyclin B (Hunt, T.
 15 (1991) Nature 349:100-101). Several types of cyclins exist (Ciechanover, A. (1994) Cell 79:13-21). Progression through G1 and S phase is driven by the G1 cyclins and their catalytic subunits, including Cdk2-cyclin A, Cdk2-cyclin E, Cdk4-cyclin D and Cdk6-cyclin D. Progression through the G2-M transition is driven by the activation of mitotic CDK-cyclin complexes such as Cdc2-cyclin A, Cdc2-cyclin B1 and Cdc2-cyclin B2 complexes (reviewed in Yang, J. and Kornbluth, S. (1999)
 20 Trends in Cell Biology 9:207-210).

Cyclins are degraded through the ubiquitin conjugation system (UCS), a major pathway for the degradation of cellular proteins in eukaryotic cells and in some bacteria. The UCS mediates the elimination of abnormal proteins and regulates the half-lives of important regulatory proteins that control cellular processes such as gene transcription and cell cycle progression. The UCS is
 25 implicated in the degradation of mitotic cyclin kinases, oncoproteins, tumor suppressor genes such as p53, viral proteins, cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins (Ciechanover, supra).

Details of the cell division cycle may vary, but the basic process consists of three principle events. The first event, interphase, involves preparations for cell division, replication of the DNA, and
 30 production of essential proteins. In the second event, mitosis, the nuclear material is divided and separates to opposite sides of the cell. The final event, cytokinesis, is division and fission of the cell cytoplasm. The sequence and timing of cell cycle transitions is under the control of the cell cycle regulation system which controls the process by positive or negative regulatory circuits at various

WO 02/097032

PCT/US02/11152

check points.

Mitosis marks the end of interphase and concludes with the onset of cytokinesis. There are four stages in mitosis, occurring in the following order: prophase, metaphase, anaphase and telophase. Prophase includes the formation of bi-polar mitotic spindles, composed of microtubules and associated proteins such as dynein, which originate from polar mitotic centers. During metaphase, the nuclear material condenses and develops kinetochore fibers which aid in its physical attachment to the mitotic spindles. The ensuing movement of the nuclear material to opposite poles along the mitotic spindles occurs during anaphase. Telophase includes the disappearance of the mitotic spindles and kinetochore fibers from the nuclear material. Mitosis depends on the interaction of numerous proteins. For example, centromere-associated proteins such as CENP-A, -B, and -C, play structural roles in kinetochore formation and assembly (Saffery, R. et al. (2000) *Human Mol. Gen.* 9: 175-185).

During the M phase of eukaryotic cell cycling, structural rearrangements occur ensuring appropriate distribution of cellular components between daughter cells. Breakdown of interphase structures into smaller subunits is common. The nuclear envelope breaks into vesicles, and nuclear lamins are disassembled. Subsequent phosphorylation of these lamins occurs and is maintained until telophase, at which time the nuclear lamina structure is reformed. cDNAs responsible for encoding M phase phosphorylation (MPPs) are components of U3 small nucleolar ribonucleoprotein (snoRNP), and relocalize to the nucleolus once mitosis is complete (Westendorf, J.M. et al. (1998) *J. Biol. Chem.* 273:437-449). U3 snoRNPs are essential mediators of RNA processing events.

Proteins involved in the regulation of cellular processes such as mitosis include the Ser/Thr-protein phosphatases type 1 (PP-1). PP-1s act by dephosphorylation of key proteins involved in the metaphase-anaphase transition. The gene PP1R7 encodes the regulatory polypeptide sds22, having at least six splice variants (Ceulemans, H. et al. (1999) *Eur. J. Biochem.* 262:36-42). Sds22 modulates the activity of the catalytic subunit of PP-1s, and enhances the PP-1-dependent dephosphorylation of mitotic substrates.

Cell cycle regulatory proteins play an important role in cell proliferation and cancer. For example, failures in the proper execution and timing of cell cycle events can lead to chromosome segregation defects resulting in aneuploidy or polyploidy. This genomic instability is characteristic of transformed cells (Luca, F.C. and Winey, M. (1998) *Mol. Biol. Cell.* 9:29-46). A recently identified protein, mMOB1, is the mammalian homolog of yeast MOB1, an essential yeast gene required for completion of mitosis and maintenance of ploidy. The mammalian mMOB1 is a member of protein complexes including protein phosphatase 2A (PP2A), and its phosphorylation appears to be regulated by PP2A (Moreno, C.S. et al. (2001) *J. Biol. Chem.* 276:24253-24260). PP2A has been implicated in

WO 02/097032

PCT/US02/11152

the development of human cancers, including lung and colon cancers and leukemias.

Cell cycle regulation involves numerous proteins interacting in a sequential manner. The eukaryotic cell cycle consists of several highly controlled events whose precise order ensures successful DNA replication and cell division. Cells maintain the order of these events by making later events dependent on the successful completion of earlier events. This dependency is enforced by cellular mechanisms called checkpoints. Examples of additional cell cycle regulatory proteins include the histone deacetylases (HDACs). HDACs are involved in cell cycle regulation, and modulate chromatin structure. Human HDAC1 has been found to interact *in vitro* with the human Hus1 gene product, whose *Schizosaccharomyces pombe* homolog has been implicated in G₂/M checkpoint control (Cai, R.L. et al. (2000) J. Biol. Chem. 275:27909-27916).

DNA damage (G₁) and DNA replication (S-phase) checkpoints arrest eukaryotic cells at the G₂/M transition. This arrest provides time for DNA repair or DNA replication to occur before entry into mitosis. Thus, the G₂/M checkpoint ensures that mitosis only occurs upon completion of DNA replication and in the absence of chromosomal damage. The Hus1 gene of *Schizosaccharomyces pombe* is a cell cycle checkpoint gene, as are the rad family of genes (e.g., rad1 and rad9) (Volkmer, E. and Karnitz, L.M. (1999) J. Biol. Chem. 274:567-570; Kostrub C.F. et al. (1998) EMBO J. 17:2055-2066). These genes are involved in the mitotic checkpoint, and are induced by either DNA damage or blockage of replication. Induction of DNA damage or replication block leads to loss of function of the Hus1 gene and subsequent cell death. Human homologs have been identified for most of the rad genes, including ATM and ATR, the human homologs of rad3p. Mutations in the ATM gene are correlated with the severe congenital disease ataxia-telangiectasia (Savitsky, K. et al. (1995) Science 268:1749-1753). The human Hus1 protein has been shown to act in a complex with rad1 protein which interacts with rad9, making them central components of a DNA damage-responsive protein complex of human cells (Volkmer, E. and Karnitz, L.M. (1999) J. Biol. Chem. 274:567-570).

The entry and exit of a cell from mitosis is regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins act by binding to and activating a group of cyclin-dependent protein kinases (Cdks) which then phosphorylate and activate selected proteins involved in the mitotic process. Cyclins are characterized by a large region of shared homology that is approximately 180 amino acids in length and referred to as the "cyclin box" (Chapman, D.L. and Wolgemuth, D.J. (1993) Development 118:229-40). In addition, cyclins contain a conserved 9 amino acid sequence in the N-terminal region of the molecule called the "destruction box". This sequence is believed to be a recognition code that triggers ubiquitin-mediated degradation of cyclin B (Hunt, T. (1991) Nature 349:100-101). Several types of cyclins exist (Ciechanover, A. (1994) Cell 79:13-21).

WO 02/097032

PCT/US02/11152

Progression through G1 and S phase is driven by the G1 cyclins and their catalytic subunits, including Cdk2-cyclin A, Cdk2-cyclin E, Cdk4-cyclin D and Cdk6-cyclin D. Progression through the G2-M transition is driven by the activation of mitotic CDK-cyclin complexes such as Cdc2-cyclin A, Cdc2-cyclin B1 and Cdc2-cyclin B2 complexes (reviewed in Yang, J. and Kornbluth, S. (1999)

5 Trends in Cell Biology 9:207-210).

Cyclins are degraded through the ubiquitin conjugation system (UCS), a major pathway for the degradation of cellular proteins in eukaryotic cells and in some bacteria. The UCS mediates the elimination of abnormal proteins and regulates the half-lives of important regulatory proteins that control cellular processes such as gene transcription and cell cycle progression. The UCS is

10 implicated in the degradation of mitotic cyclin kinases, oncoproteins, tumor suppressor genes such as p53, viral proteins, cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins (Ciechanover, supra).

The process of ubiquitin conjugation and protein degradation occurs in five principle steps (Jentsch, S. (1992) *Annu. Rev. Genet.* 26:179-207). First ubiquitin (Ub), a small, heat stable protein is

15 activated by a ubiquitin-activating enzyme (E1) in an ATP dependent reaction which binds the C-terminus of Ub to the thiol group of an internal cysteine residue in E1. Second, activated Ub is transferred to one of several Ub-conjugating enzymes (E2). Different ubiquitin-dependent proteolytic pathways employ structurally similar, but distinct ubiquitin-conjugating enzymes that are associated

20 with recognition subunits which direct them to proteins carrying a particular degradation signal. Third, E2 transfers the Ub molecule through its C-terminal glycine to a member of the ubiquitin-protein ligase family, E3. Fourth, E3 transfers the Ub molecule to the target protein. Additional Ub molecules may be added to the target protein forming a multi-Ub chain structure. Fifth, the ubiquitinated protein is then recognized and degraded by the proteasome, a large, multisubunit proteolytic enzyme complex, and Ub is released for re-utilization.

25 Prior to activation, Ub is usually expressed as a fusion protein composed of an N-terminal ubiquitin and a C-terminal extension protein (CEP) or as a polyubiquitin protein with Ub monomers attached head to tail. CEPs have characteristics of a variety of regulatory proteins; most are highly basic, contain up to 30% lysine and arginine residues, and have nucleic acid-binding domains (Monia, B.P. et al. (1989) *J. Biol. Chem.* 264:4093-4103). The fusion protein is an important intermediate

30 which appears to mediate co-regulation of the cell's translational and protein degradation activities, as well as localization of the inactive enzyme to specific cellular sites. Once delivered, C-terminal hydrolases cleave the fusion protein to release a functional Ub (Monia et al., supra).

Ub-conjugating enzymes (E2s) are important for substrate specificity in different UCS

WO 02/097032

PCT/US02/11152

pathways. All E2s have a conserved domain of approximately 16 kDa called the UBC domain that is at least 35% identical in all E2s and contains a centrally located cysteine residue required for ubiquitin-enzyme thiolester formation (Jentsch, supra). A well conserved proline-rich element is located N-terminal to the active cysteine residue. Structural variations beyond this conserved domain are used to classify the E2 enzymes. Class I E2s consist almost exclusively of the conserved UBC domain. Class II E2s have various unrelated C-terminal extensions that contribute to substrate specificity and cellular localization. Class III E2s have unique N-terminal extensions which are believed to be involved in enzyme regulation or substrate specificity.

A mitotic cyclin-specific E2 (E2-C) is characterized by the conserved UBC domain, an N-terminal extension of 30 amino acids not found in other E2s, and a 7 amino acid unique sequence adjacent to this extension. These characteristics together with the high affinity of E2-C for cyclin identify it as a new class of E2 (Aristarkhov, A. et al. (1996) Proc. Natl. Acad. Sci. 93:4294-99).

Ubiquitin-protein ligases (E3s) catalyze the last step in the ubiquitin conjugation process, covalent attachment of ubiquitin to the substrate. E3 plays a key role in determining the specificity of the process. Only a few E3s have been identified so far. One type of E3 ligases is the HECT (homologous to E6-AP C-terminus) domain protein family. One member of the family, E6-AP (E6-associated protein) is required, along with the human papillomavirus (HPV) E6 oncoprotein, for the ubiquitination and degradation of p53 (Scheffner et al. (1993) Cell 75:493-505). The C-terminal domain of HECT proteins contains the highly conserved ubiquitin-binding cysteine residue. The N-terminal region of the various HECT proteins is variable and is believed to be involved in specific substrate recognition (Huibregtse, J.M. et al. (1997) Proc. Natl Acad. Sci. USA 94:3656-3661). The SCF (Skp1-Cdc53/Cullin-F box receptor) family of proteins comprise another group of ubiquitin ligases (Deshaies, R. (1999) Annu. Rev. Dev. Biol. 15:435-467). Multiple proteins are recruited into the SCF complex, including Skp1, cullin, and an F box domain containing protein. The F box protein binds the substrate for the ubiquitination reaction and may play roles in determining substrate specificity and orienting the substrate for reaction. Skp1 interacts with both the F box protein and cullin and may be involved in positioning the F box protein and cullin in the complex for transfer of ubiquitin from the E2 enzyme to the protein substrate. Substrates of SCF ligases include proteins involved in regulation of CDK activity, activation of transcription, signal transduction, assembly of kinetochores, and DNA replication.

Sgt1 was identified in a screen for genes in yeast that suppress defects in kinetochore function caused by mutations in Skp1 (Kitagawa, K. et al. (1999) Mol. Cell 4:21-33). Sgt1 interacts with Skp1 and associates with SCF ubiquitin ligase. Defects in Sgt1 cause arrest of cells at either G1 or G2

WO 02/097032

PCT/US02/11152

stages of the cell cycle. A yeast Sgt1 null mutant can be rescued by human Sgt1, an indication of the conservation of Sgt1 function across species. Sgt1 is required for assembly of kinetochore complexes in yeast.

Abnormal activities of the UCS are implicated in a number of diseases and disorders. These include, e.g., cachexia (Llovera, M. et al. (1995) *Int. J. Cancer* 61: 138-141), degradation of the tumor-suppressor protein, p53 (Ciechanover, supra), and neurodegeneration such as observed in Alzheimer's disease (Gregori, L. et al. (1994) *Biochem. Biophys. Res. Commun.* 203: 1731-1738). Since ubiquitin conjugation is a rate-limiting step in antigen presentation, the ubiquitin degradation pathway may also have a critical role in the immune response (Grant E.P. et al. (1995) *J. Immunol.* 155: 3750-3758).

Certain cell proliferation disorders can be identified by changes in the protein complexes that normally control progression through the cell cycle. A primary treatment strategy involves reestablishing control over cell cycle progression by manipulation of the proteins involved in cell cycle regulation (Nigg, E.A. (1995) *BioEssays* 17:471-480).

Tumor necrosis factor (TNF) and related cytokines induce apoptosis in lymphoid cells. (Reviewed in Nagata, S. (1997) *Cell* 88:355-365.) Binding of TNF to its receptor triggers a signal transduction pathway that results in the activation of a cascade of related proteases, called caspases. One such caspase, ICE (Interleukin-1 β converting enzyme), is a cysteine protease comprised of two large and two small subunits generated by ICE auto-cleavage. (Dinarello, C. A. (1994) *FASEB J.* 8:1314-1325.) ICE is expressed primarily in monocytes. ICE processes the cytokine precursor, interleukin-1 β , into its active form, which plays a central role in acute and chronic inflammation, bone resorption, myelogenous leukemia, and other pathological processes. ICE and related caspases cause apoptosis when overexpressed in transfected cell lines.

A final step in the apoptotic effector pathway is the fragmentation of nuclear DNA. Recently, a novel factor linking caspase activity to DNA fragmentation has been identified. (Xuesong, L. et al. (1997) *Cell* 89:175-184.) This factor, DNA fragmentation factor 45 (DFF-45), is proteolytically activated by caspase and is required for DNA fragmentation. DFF-45 is 331 amino acids in length and exists in the cell as a heterodimer with a second uncharacterized factor. The amino acid sequence of DFF-45 indicates that it is not a nuclease, suggesting that DFF-45 may activate a downstream nuclease. In addition, mRNA encoding a protein related to DFF-45 has been isolated from mouse adipogenic cells. (Danesch, U. et al. (1992) *J. Biol. Chem.* 267:7185-7193.) Expression of this mRNA is induced in steroid-treated, differentiating adipocytes. The predicted protein, FSP-27 (fat cell-specific, 27 kilodaltons), is highly basic with a predicted isoelectric point of 10.

WO 02/097032

PCT/US02/11152

Dysregulation of apoptosis has recently been recognized as a significant factor in the pathogenesis of many human diseases. For example, excessive cell survival caused by decreased apoptosis can contribute to disorders related to cell proliferation and the immune response. Such disorders include cancer, autoimmune diseases, viral infections, and inflammation. In contrast, excessive cell death caused by increased apoptosis can lead to degenerative and immunodeficiency disorders such as AIDS, neurodegenerative diseases, and myelodysplastic syndromes. (Thompson, C.B. (1995) Science 267:1456-1462.)

Embryogenesis

Mammalian embryogenesis is a process which encompasses the first few weeks of development following conception. During this period, embryogenesis proceeds from a single fertilized egg to the formation of the three embryonic tissues, then to an embryo which has most of its internal organs and all of its external features.

The normal course of mammalian embryogenesis depends on the correct temporal and spatial regulation of a large number of genes and tissues. These regulation processes have been intensely studied in mouse. An essential process that is still poorly understood is the activation of the embryonic genome after fertilization. As mouse oocytes grow, they accumulate transcripts that are either translated directly into proteins or stored for later activation by regulated polyadenylation. During subsequent meiotic maturation and ovulation, the maternal genome is transcriptionally inert, and most maternal transcripts are deadenylated and/or degraded prior to, or together with, the activation of the zygotic genes at the two-cell stage (Stutz, A. et al. (1998) Genes Dev. 12:2535-2548). The maternal to embryonic transition involves the degradation of oocyte, but not zygotic transcripts, the activation of the embryonic genome, and the induction of cell cycle progression to accommodate early development.

MATER (Maternal Antigen That Embryos Require) was initially identified as a target of antibodies from mice with ovarian immunity (Tong, Z.-B., and Nelson, L.M. (1999) Endocrinology 140:3720-3726). Expression of the gene encoding MATER is restricted to the oocyte, making it one of a limited number of known maternal-effect genes in mammals (Tong, Z.-B., et al. (2000) Mamm. Genome 11:281-287). The MATER protein is required for embryonic development beyond two cells, based upon preliminary results from mice in which this gene has been inactivated. The 1111-amino acid MATER protein contains a hydrophilic repeat region in the amino terminus, and a region containing 14 leucine-rich repeats in the carboxyl terminus. These repeats resemble the sequence found in porcine ribonuclease inhibitor that is critical for protein-protein interactions.

The degradation of maternal transcripts during meiotic maturation and ovulation may involve

WO 02/097032

PCT/US02/11152

the activation of a ribonuclease just prior to ovulation. Thus the function of MATER may be to bind to the maternal ribonuclease and prevent degradation of zygotic transcripts (Tong (2000) supra). In addition to its role in oocyte development and embryogenesis, MATER may also be relevant to the pathogenesis of ovarian immunity, as it is a target of autoantibodies in mice with autoimmune oophoritis (Tong (1999) supra).

The maternal mRNA D7 is a moderately abundant transcript in *Xenopus laevis* whose expression is highest in, and perhaps restricted to, oogenesis and early embryogenesis. The D7 protein is absent from oocytes and first begins to accumulate during oocyte maturation. Its levels are highest during the first day of embryonic development and then they decrease. The loss of D7 protein affects the maturation process itself, significantly delaying the time course of germinal vesicle breakdown. Thus, D7 is a newly described protein involved in oocyte maturation (Smith R.C., et al. (1988) *Genes Dev.* 2(10):1296-306.)

Many other genes are involved in subsequent stages of embryogenesis. After fertilization, the oocyte is guided by fimbria at the distal end of each fallopian tube into and through the fallopian tube and thence into the uterus. Changes in the uterine endometrium prepare the tissue to support the implantation and embryonic development of a fertilized ovum. Several stages of division have occurred before the dividing ovum, now a blastocyst with about 100 cells, enters the uterus. Upon reaching the uterus, the developing blastocyst usually remains in the uterine cavity an additional two to four days before implanting in the endometrium, the inner lining of the uterus. Implantation results from the action of trophoblast cells that develop over the surface of the blastocyst. These cells secrete proteolytic enzymes that digest and liquefy the cells of the endometrium. The invasive process is reviewed in Fisher and Damsky (1993; *Semin Cell Biol* 4:183-188) and Graham and Lala (1992; *Biochem Cell Biol* 70:867-874). Once implantation has taken place, the trophoblast and other sublying cells proliferate rapidly, forming the placenta and the various membranes of pregnancy. (See Guyton, A.C. (1991) *Textbook of Medical Physiology*, 8th ed. W.B. Saunders Company, Philadelphia pp. 915-919.)

The placenta has an essential role in protecting and nourishing the developing fetus. In most species the syncytiotrophoblast layer is present on the outside of the placenta at the fetal-maternal interface. This is a continuous structure, one cell deep, formed by the fusion of the constituent trophoblast cells. The syncytiotrophoblast cells play important roles in maternal-fetal exchange, in tissue remodeling during fetal development, and in protecting the developing fetus from the maternal immune response (Stoye, J.P. and Coffin, J.M. (2000) *Nature* 403:715-717).

A gene called syncytin is the envelope gene of a human endogenous defective provirus.

WO 02/097032

PCT/US02/11152

Syncytin is expressed in high levels in placenta, and more weakly in testis, but is not detected in any other tissues (Mi, S. et al. (2000) Nature 403:785-789). Syncytin expression in the placenta is restricted to the syncytiotrophoblasts. Since retroviral env proteins are often involved in promoting cell fusion events, it was thought that syncytin might be involved in regulating the fusion of trophoblast cells into the syncytiotrophoblast layer. Experiments demonstrated that syncytin can mediate cell fusion in vitro, and that anti-syncytin antibodies can inhibit the fusion of placental cytotrophoblasts (Mi, supra). In addition, a conserved immunosuppressive domain present in retroviral envelope proteins, and found in syncytin at amino acid residues 373-397, might be involved in preventing maternal immune responses against the developing embryo.

Syncytin may also be involved in regulating trophoblast invasiveness by inducing trophoblast fusion and terminal differentiation (Mi, supra). Insufficient trophoblast infiltration of the uterine wall is associated with placental disorders such as preeclampsia, or pregnancy induced hypertension, while uncontrolled trophoblast invasion is observed in choriocarcinoma and other gestational trophoblastic diseases. Thus syncytin function may be involved in these diseases.

15 Cell Division

Cell division is the fundamental process by which all living things grow and reproduce. In unicellular organisms such as yeast and bacteria, each cell division doubles the number of organisms, while in multicellular species many rounds of cell division are required to replace cells lost by wear or by programmed cell death, and for cell differentiation to produce a new tissue or organ. Details of the cell division cycle may vary, but the basic process consists of three principle events. The first event, interphase, involves preparations for cell division, replication of the DNA, and production of essential proteins. In the second event, mitosis, the nuclear material is divided and separates to opposite sides of the cell. The final event, cytokinesis, is division and fission of the cell cytoplasm. The sequence and timing of cell cycle transitions is under the control of the cell cycle regulation system which controls the process by positive or negative regulatory circuits at various check points.

Regulated progression of the cell cycle depends on the integration of growth control pathways with the basic cell cycle machinery. Cell cycle regulators have been identified by selecting for human and yeast cDNAs that block or activate cell cycle arrest signals in the yeast mating pheromone pathway when they are overexpressed. Known regulators include human CPR (cell cycle progression restoration) genes, such as CPR8 and CPR2, and yeast CDC (cell division control) genes, including CDC91, that block the arrest signals. The CPR genes express a variety of proteins including cyclins, tumor suppressor binding proteins, chaperones, transcription factors, translation factors, and RNA-binding proteins (Edwards, M.C. et al.(1997) Genetics 147:1063-1076).

WO 02/097032

PCT/US02/11152

Several cell cycle transitions, including the entry and exit of a cell from mitosis, are dependent upon the activation and inhibition of cyclin-dependent kinases (Cdks). The Cdks are composed of a kinase subunit, Cdk, and an activating subunit, cyclin, in a complex that is subject to many levels of regulation. There appears to be a single Cdk in *Saccharomyces cerevisiae* and *Saccharomyces pombe* whereas mammals have a variety of specialized Cdks. Cyclins act by binding to and activating cyclin-dependent protein kinases which then phosphorylate and activate selected proteins involved in the mitotic process. The Cdk-cyclin complex is both positively and negatively regulated by phosphorylation, and by targeted degradation involving molecules such as CDC4 and CDC53. In addition, Cdks are further regulated by binding to inhibitors and other proteins such as Sic1 that modify their specificity or accessibility to regulators (Patra, D. and W.G. Dunphy (1996) *Genes Dev.* 10:1503-1515; and Mathias, N. et al. (1996) *Mol. Cell Biol.* 16:6634-6643).

Reproduction

The male and female reproductive systems are complex and involve many aspects of growth and development. The anatomy and physiology of the male and female reproductive systems are reviewed in (Guyton, A.C. (1991) *Textbook of Medical Physiology*, W.B. Saunders Co., Philadelphia PA, pp. 899-928).

The male reproductive system includes the process of spermatogenesis, in which the sperm are formed, and male reproductive functions are regulated by various hormones and their effects on accessory sexual organs, cellular metabolism, growth, and other bodily functions.

Spermatogenesis begins at puberty as a result of stimulation by gonadotropic hormones released from the anterior pituitary. Immature sperm (spermatogonia) undergo several mitotic cell divisions before undergoing meiosis and full maturation. The testes secrete several male sex hormones, the most abundant being testosterone, that is essential for growth and division of the immature sperm, and for the masculine characteristics of the male body. Three other male sex hormones, gonadotropin-releasing hormone (GnRH), luteinizing hormone (LH), and follicle-stimulating hormone (FSH) control sexual function.

The uterus, ovaries, fallopian tubes, vagina, and breasts comprise the female reproductive system. The ovaries and uterus are the source of ova and the location of fetal development, respectively. The fallopian tubes and vagina are accessory organs attached to the top and bottom of the uterus, respectively. Both the uterus and ovaries have additional roles in the development and loss of reproductive capability during a female's lifetime. The primary role of the breasts is lactation. Multiple endocrine signals from the ovaries, uterus, pituitary, hypothalamus, adrenal glands, and other tissues coordinate reproduction and lactation. These signals vary during the monthly menstruation

WO 02/097032

PCT/US02/11152

cycle and during the female's lifetime. Similarly, the sensitivity of reproductive organs to these endocrine signals varies during the female's lifetime.

A combination of positive and negative feedback to the ovaries, pituitary and hypothalamus glands controls physiologic changes during the monthly ovulation and endometrial cycles. The anterior pituitary secretes two major gonadotropin hormones, follicle-stimulating hormone (FSH) and luteinizing hormone (LH), regulated by negative feedback of steroids, most notably by ovarian estradiol. If fertilization does not occur, estrogen and progesterone levels decrease. This sudden reduction of the ovarian hormones leads to menstruation, the desquamation of the endometrium.

Hormones further govern all the steps of pregnancy, parturition, lactation, and menopause.

During pregnancy large quantities of human chorionic gonadotropin (hCG), estrogens, progesterone, and human chorionic somatomammotropin (hCS) are formed by the placenta. hCG, a glycoprotein similar to luteinizing hormone, stimulates the corpus luteum to continue producing more progesterone and estrogens, rather than to involute as occurs if the ovum is not fertilized. hCS is similar to growth hormone and is crucial for fetal nutrition.

The female breast also matures during pregnancy. Large amounts of estrogen secreted by the placenta trigger growth and branching of the breast milk ductal system while lactation is initiated by the secretion of prolactin by the pituitary gland.

Parturition involves several hormonal changes that increase uterine contractility toward the end of pregnancy, as follows. The levels of estrogens increase more than those of progesterone.

Oxytocin is secreted by the neurohypophysis. Concomitantly, uterine sensitivity to oxytocin increases. The fetus itself secretes oxytocin, cortisol (from adrenal glands), and prostaglandins.

Menopause occurs when most of the ovarian follicles have degenerated. The ovary then produces less estradiol, reducing the negative feedback on the pituitary and hypothalamus glands. Mean levels of circulating FSH and LH increase, even as ovulatory cycles continue. Therefore, the ovary is less responsive to gonadotropins, and there is an increase in the time between menstrual cycles. Consequently, menstrual bleeding ceases and reproductive capability ends.

Cell Differentiation and Proliferation

Tissue growth involves complex and ordered patterns of cell proliferation, cell differentiation, and apoptosis. Cell proliferation must be regulated to maintain both the number of cells and their spatial organization. This regulation depends upon the appropriate expression of proteins which control cell cycle progression in response to extracellular signals, such as growth factors and other mitogens, and intracellular cues, such as DNA damage or nutrient starvation. Molecules which directly or indirectly modulate cell cycle progression fall into several categories, including growth factors and

WO 02/097032

PCT/US02/11152

their receptors, second messenger and signal transduction proteins, oncogene products, tumor-suppressor proteins, and mitosis-promoting factors.

Growth factors were originally described as serum factors required to promote cell proliferation. Most growth factors are large, secreted polypeptides that act on cells in their local environment. Growth factors bind to and activate specific cell surface receptors and initiate intracellular signal transduction cascades. Many growth factor receptors are classified as receptor tyrosine kinases which undergo autophosphorylation upon ligand binding. Autophosphorylation enables the receptor to interact with signal transduction proteins characterized by the presence of SH2 or SH3 domains (Src homology regions 2 or 3). These proteins then modulate the activity state of small G-proteins, such as Ras, Rab, and Rho, along with GTPase activating proteins (GAPs), guanine nucleotide releasing proteins (GNRPs), and other guanine nucleotide exchange factors. Small G proteins act as molecular switches that activate other downstream events, such as mitogen-activated protein kinase (MAP kinase) cascades. MAP kinases ultimately activate transcription of mitosis-promoting genes.

In addition to growth factors, small signaling peptides and hormones also influence cell proliferation. These molecules bind primarily to another class of receptor, the trimeric G-protein coupled receptor (GPCR), found predominantly on the surface of immune, neuronal and neuroendocrine cells. Upon ligand binding, the GPCR activates a trimeric G protein which in turn triggers increased levels of intracellular second messengers such as phospholipase C, Ca^{2+} , and cyclic AMP. Most GPCR-mediated signaling pathways indirectly promote cell proliferation by causing the secretion or breakdown of other signaling molecules that have direct mitogenic effects. These signaling cascades often involve activation of kinases and phosphatases. Some growth factors, such as some members of the transforming growth factor beta (TGF- β) family, act on some cells to stimulate cell proliferation and on other cells to inhibit it. Growth factors may also stimulate a cell at one concentration and inhibit the same cell at another concentration. Most growth factors also have a multitude of other actions besides the regulation of cell growth and division: they can control the proliferation, survival, differentiation, migration, or function of cells depending on the circumstance. For example, the tumor necrosis factor/nerve growth factor (TNF/NGF) family can activate or inhibit cell death, as well as regulate proliferation and differentiation. The cell response depends on the type of cell, its stage of differentiation and transformation status, which surface receptors are stimulated, and the types of stimuli acting on the cell (Smith, A. et al. (1994) Cell 76:959-962; and Nocentini, G. et al. (1997) Proc. Natl. Acad. Sci. USA 94:6216-6221).

Neighboring cells in a tissue compete for growth factors, and when provided with "unlimited"

WO 02/097032

PCT/US02/11152

quantities in a perfused system will grow to even higher cell densities before reaching density-dependent inhibition of cell division. Cells often demonstrate an anchorage dependence of cell division as well. This anchorage dependence may be associated with the formation of focal contacts linking the cytoskeleton with the extracellular matrix (ECM). The expression of ECM components can be stimulated by growth factors. For example, TGF- β stimulates fibroblasts to produce a variety of ECM proteins, including fibronectin, collagen, and tenascin (Pearson, C.A. et al. (1988) EMBO J. 7:2677-2981). In fact, for some cell types specific ECM molecules, such as laminin or fibronectin, may act as growth factors. Tenascin-C and -R, expressed in developing and lesioned neural tissue, provide stimulatory/anti-adhesive or inhibitory properties, respectively, for axonal growth (Faissner, A. (1997) Cell Tissue Res. 290:331-341).

Cancers are associated with the activation of oncogenes which are derived from normal cellular genes. These oncogenes encode oncoproteins which convert normal cells into malignant cells. Some oncoproteins are mutant isoforms of the normal protein, and other oncoproteins are abnormally expressed with respect to location or amount of expression. The latter category of oncoprotein causes cancer by altering transcriptional control of cell proliferation. Five classes of oncoproteins are known to affect cell cycle controls. These classes include growth factors, growth factor receptors, intracellular signal transducers, nuclear transcription factors, and cell-cycle control proteins. Viral oncogenes are integrated into the human genome after infection of human cells by certain viruses. Examples of viral oncogenes include v-src, v-abl, and v-fps. Many cases related to the overexpression of proteins associated with tumors and metastasis have been reported. The Mta1 gene has been cloned in mice, in both cell lines and tissues representing metastatic tumors (Simpson, A. et al. (2001) Gene 273:29-39). Expression of the melanoma antigen-encoding gene (MAGE) family of proteins has also been detected in many tumors. GAC1, a new member of the leucine-rich repeat superfamily, is amplified and overexpressed in malignant gliomas (Almeida, A. et al. (1998) Oncogene 16:2997-3002).

Many oncogenes have been identified and characterized. These include sis, erbA, erbB, her-2, mutated G, src, abl, ras, crk, jun, fos, myc, and mutated tumor-suppressor genes such as RB, p53, mdm2, Cip1, p16, and cyclin D. Transformation of normal genes to oncogenes may also occur by chromosomal translocation. The Philadelphia chromosome, characteristic of chronic myeloid leukemia and a subset of acute lymphoblastic leukemias, results from a reciprocal translocation between chromosomes 9 and 22 that moves a truncated portion of the proto-oncogene c-abl to the breakpoint cluster region (bcr) on chromosome 22.

Tumor-suppressor genes are involved in regulating cell proliferation. Mutations which cause

WO 02/097032

PCT/US02/11152

reduced or loss of function in tumor-suppressor genes result in uncontrolled cell proliferation. For example, the retinoblastoma gene product (RB), in a non-phosphorylated state, binds several early-response genes and suppresses their transcription, thus blocking cell division. Phosphorylation of RB causes it to dissociate from the genes, releasing the suppression, and allowing cell division to proceed.

5 SEB (SET-binding protein) is a novel nuclear protein that interacts in a yeast two-hybrid system and in human cells with SET, the translocation breakpoint-encoded protein in acute undifferentiated leukemia. SEB also has an oncoprotein Ski homologous region, six PEST sequences and three sequential PPLPPPPP repeats at the C-terminus. SEB mRNA is expressed ubiquitously in all examined human adult tissues and cells. SET has been mapped to chromosome 18q21.1. This
10 reagon also contains tumor suppressor genes associated with deletions in cancer and leukemia (Minakuchi, M. et al. (2001) Eru. J. Biochem. 268:1340-1351).

Cell Differentiation

Multicellular organisms are comprised of diverse cell types that differ dramatically both in structure and function, despite the fact that each cell is like the others in its hereditary endowment.
15 Cell differentiation is the process by which cells come to differ in their structure and physiological function. The cells of a multicellular organism all arise from mitotic divisions of a single-celled zygote. The zygote is totipotent, meaning that it has the ability to give rise to every type of cell in the adult body. During development the cellular descendants of the zygote lose their totipotency and become determined. Once its prospective fate is achieved, a cell is said to have differentiated. All
20 descendants of this cell will be of the same type.

Human growth and development requires the spatial and temporal regulation of cell differentiation, along with cell proliferation and regulated cell death. These processes coordinate to control reproduction, aging, embryogenesis, morphogenesis, organogenesis, and tissue repair and maintenance. The processes involved in cell differentiation are also relevant to disease states such as
25 cancer, in which case the factors regulating normal cell differentiation have been altered, allowing the cancerous cells to proliferate in an anaplastic, or undifferentiated, state.

The mechanisms of differentiation involve cell-specific regulation of transcription and translation, so that different genes are selectively expressed at different times in different cells. Genetic experiments using the fruit fly *Drosophila melanogaster* have identified regulated cascades of
30 transcription factors which control pattern formation during development and differentiation. These include the homeotic genes, which encode transcription factors containing homeobox motifs. The products of homeotic genes determine how the insect's imaginal discs develop from masses of undifferentiated cells to specific segments containing complex organs. Many genes found to be

WO 02/097032

PCT/US02/11152

involved in cell differentiation and development in Drosophila have homologs in mammals. Some human genes have equivalent developmental roles to their Drosophila homologs. The human homolog of the Drosophila eyes absent gene (*eya*) underlies branchio-oto-renal syndrome, a developmental disorder affecting the ears and kidneys (Abdelhak, S. et al. (1997) Nat. Genet. 15:157-164). The

5 Drosophila slit gene encodes a secreted leucine-rich repeat containing protein expressed by the midline glial cells and required for normal neural development.

At the cellular level, growth and development are governed by the cell's decision to enter into or exit from the cell cycle and by the cell's commitment to a terminally differentiated state.

Differential gene expression within cells is triggered in response to extracellular signals and other
10 environmental cues. Such signals include growth factors and other mitogens such as retinoic acid; cell-cell and cell-matrix contacts; and environmental factors such as nutritional signals, toxic substances, and heat shock. Candidate genes that may play a role in differentiation can be identified by altered expression patterns upon induction of cell differentiation in vitro.

The final step in cell differentiation results in a specialization that is characterized by the
15 production of particular proteins, such as contractile proteins in muscle cells, serum proteins in liver cells and globins in red blood cell precursors. The expression of these specialized proteins depends at least in part on cell-specific transcription factors. For example, the homobox-containing transcription factor PAX-6 is essential for early eye determination, specification of ocular tissues, and normal eye development in vertebrates.

In the case of epidermal differentiation, the induction of differentiation-specific genes occurs
20 either together with or following growth arrest and is believed to be linked to the molecular events that control irreversible growth arrest. Irreversible growth arrest is an early event which occurs when cells transit from the basal to the innermost suprabasal layer of the skin and begin expressing squamous-specific genes. These genes include those involved in the formation of the cross-linked
25 envelope, such as transglutaminase I and III, involucrin, loricin, and small proline-rich repeat (SPRR) proteins. The SPRR proteins are 8-10 kDa in molecular mass, rich in proline, glutamine, and cysteine, and contain similar repeating sequence elements. The SPRR proteins may be structural proteins with a strong secondary structure or metal-binding proteins such as metallothioneins. (Jettten, A. M. and Harvat, B. L. (1997) J. Dermatol. 24:711-725; PRINTS Entry PR00021 PRORICH Small proline-rich
30 protein signature.)

The Wnt gene family of secreted signaling molecules is highly conserved throughout eukaryotic cells. Members of the Wnt family are involved in regulating chondrocyte differentiation within the cartilage template. Wnt-5a, Wnt-5b and Wnt-4 genes are expressed in chondrogenic

WO 02/097032

PCT/US02/11152

regions of the chicken limb, Wnt-5a being expressed in the perichondrium (mesenchymal cells immediately surrounding the early cartilage template). Wnt-5a misexpression delays the maturation of chondrocytes and the onset of bone collar formation in chicken limb (Hartmann, C. and Tabin, C.J. (2000) *Development* 127:3141-3159).

5 Glypicans are a family of cell surface heparan sulfate proteoglycans that play an important role in cellular growth control and differentiation. Cerebroglycan, a heparan sulfate proteoglycan expressed in the nervous system, is involved with the motile behavior of developing neurons (Stipp, C.S. et al. (1994) *J. Cell Biol.* 124:149-160).

Notch plays an active role in the differentiation of glial cells, and influences the length and
10 organization of neuronal processes (for a review, see Frisen, J. and Lendahl, U. (2001) *Bioessays* 23:3-7). The Notch receptor signaling pathway is important for morphogenesis and development of many organs and tissues in multicellular species. *Drosophila* fringe proteins modulate the activation of the Notch signal transduction pathway at the dorsal-ventral boundary of the wing imaginal disc. Mammalian fringe-related family members participate in boundary determination during segmentation
15 (Johnston, S.H. et al. (1997) *Development* 124:2245-2254).

Recently a number of proteins have been found to contain a conserved cysteine-rich domain of about 60 amino-acid residues called the LIM domain (for Lin-11 Isl-1 Mec-3) (Freyd G. et al. (1990) *Nature* 344:876-879; Baltz R. et al. (1992) *Plant Cell* 4:1465-1466). In the LIM domain, there are seven conserved cysteine residues and a histidine. The LIM domain binds two zinc ions
20 (Michelsen J.W. et al. (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:4404-4408). LIM does not bind DNA, rather it seems to act as an interface for protein-protein interaction.

Apoptosis

Normal development, growth, and homeostasis in multicellular organisms require a careful balance between the production and destruction of cells in tissues throughout the body. Cell division is
25 a carefully coordinated process with numerous checkpoints and control mechanisms. These mechanisms are designed to regulate DNA replication and to prevent inappropriate or excessive cell proliferation. In contrast, apoptosis is the genetically controlled process by which unneeded or defective cells undergo programmed cell death. Unlike necrotic or injured cells, apoptotic cells are rapidly phagocytosed by neighboring cells or macrophages without leaking their potentially damaging
30 contents into the surrounding tissue or triggering an inflammatory response.

Apoptosis is the genetically controlled process by which unneeded or defective cells undergo programmed cell death. Selective elimination of cells is as important for morphogenesis and tissue remodeling as is cell proliferation and differentiation. Lack of apoptosis may result in hyperplasia and

WO 02/097032

PCT/US02/11152

other disorders associated with increased cell proliferation. Apoptosis is also a critical component of the immune response. Immune cells such as cytotoxic T-cells and natural killer cells prevent the spread of disease by inducing apoptosis in tumor cells and virus-infected cells. In addition, immune cells that fail to distinguish self molecules from foreign molecules must be eliminated by apoptosis to

5 avoid an autoimmune response.

Apoptotic cells undergo distinct morphological changes. Hallmarks of apoptosis include cell shrinkage, nuclear and cytoplasmic condensation, and alterations in plasma membrane topology. Biochemically, apoptotic cells are characterized by increased intracellular calcium concentration, fragmentation of chromosomal DNA, and expression of novel cell surface components.

10 The molecular mechanisms of apoptosis are highly conserved, and many of the key protein regulators and effectors of apoptosis have been identified. Apoptosis generally proceeds in response to a signal which is transduced intracellularly and results in altered patterns of gene expression and protein activity. Signaling molecules such as hormones and cytokines are known both to stimulate and to inhibit apoptosis through interactions with cell surface receptors. Transcription factors also play an

15 important role in the onset of apoptosis. A number of downstream effector molecules, especially proteases, have been implicated in the degradation of cellular components and the proteolytic activation of other apoptotic effectors.

The Bcl-2 family of proteins, as well as other cytoplasmic proteins, are key regulators of apoptosis. There are at least 15 Bcl-2 family members within 3 subfamilies. These proteins have

20 been identified in mammalian cells and in viruses, and each possesses at least one of four Bcl-2 homology domains (BH1 to BH4), which are highly conserved. Bcl-2 family proteins contain the BH1 and BH2 domains, which are found in members of the pro-survival subfamily, while those proteins which are most similar to Bcl-2 have all four conserved domains, enabling inhibition of apoptosis following encounters with a variety of cytotoxic challenges. Members of the pro-survival subfamily

25 include Bcl-2, Bcl-x_L, Bcl-w, Mcl-1, and A1 in mammals; NF-13 (chicken); CED-9 (Caenorhabditis elegans); and viral proteins BHRF1, LMW5-HL, ORF16, KS-Bcl-2, and E1B-19K. The BH3 domain is essential for the function of pro-apoptosis subfamily proteins. The two pro-apoptosis subfamilies, Bax and BH3, include Bax, Bak, and Bok (also called Mtd); and Bik, Blk, Hrk, BNP3, Bim_L, Bad, Bid, and Egl-1 (C. elegans); respectively. Members of the Bax subfamily contain the BH1, BH2, and

30 BH3 domains, and resemble Bcl-2 rather closely. In contrast, members of the BH3 subfamily have only the 9-16 residue BH3 domain, being otherwise unrelated to any known protein, and only Bik and Blk share sequence similarity. The proteins of the two pro-apoptosis subfamilies may be the antagonists of pro-survival subfamily proteins. This is illustrated in C. elegans where Egl-1, which is

WO 02/097032

PCT/US02/11152

required for apoptosis, binds to and acts via CED-9 (for review, see Adams, J.M. and Cory, S. (1998) Science 281:1322-1326).

Heterodimerization between pro-apoptosis and anti-apoptosis subfamily proteins seems to have a titrating effect on the functions of these protein subfamilies, which suggests that relative concentrations of the members of each subfamily may act to regulate apoptosis. Heterodimerization is not required for a pro-survival protein; however, it is essential in the BH3 subfamily, and less so in the Bax subfamily.

The Bcl-2 protein has 2 isoforms, alpha and beta, which are formed by alternative splicing. It forms homodimers and heterodimers with Bax and Bak proteins and the Bcl-X isoform Bcl-x_s. Heterodimerization with Bax requires intact BH1 and BH2 domains, and is necessary for pro-survival activity. The BH4 domain seems to be involved in pro-survival activity as well. Bcl-2 is located within the inner and outer mitochondrial membranes, as well as within the nuclear envelope and endoplasmic reticulum, and is expressed in a variety of tissues. Its involvement in follicular lymphoma (type II chronic lymphatic leukemia) is seen in a chromosomal translocation T(14;18) (q32;q21) and involves immunoglobulin gene regions.

The Bcl-x protein is a dominant regulator of apoptotic cell death. Alternative splicing results in three isoforms, Bcl-xB, a long isoform, and a short isoform. The long isoform exhibits cell death repressor activity, while the short isoform promotes apoptosis. Bcl-xL forms heterodimers with Bax and Bak, although heterodimerization with Bax does not seem to be necessary for pro-survival (anti-apoptosis) activity. Bcl-xS forms heterodimers with Bcl-2. Bcl-x is found in mitochondrial membranes and the perinuclear envelope. Bcl-xS is expressed at high levels in developing lymphocytes and other cells undergoing a high rate of turnover. Bcl-xL is found in adult brain and in other tissues' long-lived post-mitotic cells. As with Bcl-2, the BH1, BH2, and BH4 domains are involved in pro-survival activity.

The Bcl-w protein is found within the cytoplasm of almost all myeloid cell lines and in numerous tissues, with the highest levels of expression in brain, colon, and salivary gland. This protein is expressed in low levels in testis, liver, heart, stomach, skeletal muscle, and placenta, and a few lymphoid cell lines. Bcl-w contains the BH1, BH2, and BH4 domains, all of which are needed for its cell survival promotion activity. Although mice in which Bcl-w gene function was disrupted by homologous recombination were viable, healthy, and normal in appearance, and adult females had normal reproductive function, the adult males were infertile. In these males, the initial, prepuberty stage of spermatogenesis was largely unaffected and the testes developed normally. However, the seminiferous tubules were disorganized, contained numerous apoptotic cells, and were incapable of

WO 02/097032

PCT/US02/11152

producing mature sperm. This mouse model may be applicable to some cases of human male sterility and suggests that alteration of programmed cell death in the testes may be useful in modulating fertility (Print, C.G. et al. (1998) Proc. Natl. Acad. Sci. USA 95:12424-12431).

Studies in rat ischemic brain found Bcl-w to be overexpressed relative to its normal low
 5 constitutive level of expression in nonischemic brain. Furthermore, in vitro studies to examine the mechanism of action of Bcl-w revealed that isolated rat brain mitochondria were unable to respond to an addition of recombinant Bax or high concentrations of calcium when Bcl-w was also present. The normal response would be the release of cytochrome c from the mitochondria. Additionally, recombinant Bcl-w protein was found to inhibit calcium-induced loss of mitochondrial transmembrane
 10 potential, which is indicative of permeability transition. Together these findings suggest that Bcl-w may be a neuro-protectant against ischemic neuronal death and may achieve this protection via the mitochondrial death-regulatory pathway (Yan, C. et al. (2000) J. Cereb. Blood Flow Metab. 20:620-630).

The bfl-1 gene is an additional member of the Bcl-2 family, and is also a suppressor of
 15 apoptosis. The Bfl-1 protein has 175 amino acids, and contains the BH1, BH2, and BH3 conserved domains found in Bcl-2 family members. It also contains a Gln-rich NH2-terminal region and lacks an NH domain 1, unlike other Bcl-2 family members. The mouse A1 protein shares high sequence homology with Bfl-1 and has the 3 conserved domains found in Bfl-1. Apoptosis induced by the p53 tumor suppressor protein is suppressed by Bfl-1, similar to the action of Bcl-2, Bcl-xL, and EBV-
 20 BHRF1 (D'Sa-Eipper, C. et al. (1996) Cancer Res. 56:3879-3882). Bfl-1 is found intracellularly, with the highest expression in the hematopoietic compartment, i.e. blood, spleen, and bone marrow; moderate expression in lung, small intestine, and testis; and minimal expression in other tissues. It is also found in vascular smooth muscle cells and hematopoietic malignancies. A correlation has been noted between the expression level of bfl-1 and the development of stomach cancer, suggesting that
 25 the Bfl-1 protein is involved in the development of stomach cancer, either in the promotion of cancerous cell survival or in cancer (Choi, S.S. et al. (1995) Oncogene 11:1693-1698).

Cancers are characterized by continuous or uncontrolled cell proliferation. Some cancers are associated with suppression of normal apoptotic cell death. Strategies for treatment may involve either reestablishing control over cell cycle progression, or selectively stimulating apoptosis in
 30 cancerous cells (Nigg, E.A. (1995) BioEssays 17:471-480). Immunological defenses against cancer include induction of apoptosis in mutant cells by tumor suppressors, and the recognition of tumor antigens by T lymphocytes. Response to mitogenic stresses is frequently controlled at the level of transcription and is coordinated by various transcription factors. For example, the Rel/NF-kappa B

WO 02/097032

PCT/US02/11152

family of vertebrate transcription factors plays a pivotal role in inflammatory and immune responses to radiation. The NF-kappa B family includes p50, p52, RelA, RelB, cRel, and other DNA-binding proteins. The p52 protein induces apoptosis, upregulates the transcription factor c-Jun, and activates c-Jun N-terminal kinase 1 (JNK1) (Sun, L. et al. (1998) *Gene* 208:157-166). Most NF-kappa B proteins form DNA-binding homodimers or heterodimers. Dimerization of many transcription factors is mediated by a conserved sequence known as the bZIP domain, characterized by a basic region followed by a leucine zipper.

The Fas/Apo-1 receptor (FAS) is a member of the tumor necrosis factor (TNF) receptor family. Upon binding its ligand (Fas ligand), the membrane-spanning FAS induces apoptosis by recruiting several cytoplasmic proteins that transmit the death signal. One such protein, termed FAS-associated protein factor 1 (FAF1), was isolated from mice, and it was demonstrated that expression of FAF1 in L cells potentiated FAS-induced apoptosis (Chu, K. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:11894-11898). Subsequently, FAS-associated factors have been isolated from numerous other species, including fruit fly and quail (Frohlich, T. et al. (1998) *J. Cell Sci.* 111:2353-2363). Another cytoplasmic protein that functions in the transmittal of the death signal from Fas is the Fas-associated death domain protein, also known as FADD. FADD transmits the death signal in both FAS-mediated and TNF receptor-mediated apoptotic pathways by activating caspase-8 (Bang, S. et al. (2000) *J. Biol. Chem.* 275:36217-36222).

Fragmentation of chromosomal DNA is one of the hallmarks of apoptosis. DNA fragmentation factor (DFF) is a protein composed of two subunits, a 40-kDa caspase-activated nuclease termed DFF40/CAD, and its 45-kDa inhibitor DFF45/ICAD. Two mouse homologs of DFF45/ICAD, termed CIDE-A and CIDE-B, have recently been described (Inohara, N. et al. (1998) *EMBO J.* 17:2526-2533). CIDE-A and CIDE-B expression in mammalian cells activated apoptosis, while expression of CIDE-A alone induced DNA fragmentation. In addition, FAS-mediated apoptosis was enhanced by CIDE-A and CIDE-B, further implicating these proteins as effectors that mediate apoptosis.

Transcription factors play an important role in the onset of apoptosis. A number of downstream effector molecules, particularly proteases such as the cysteine proteases called caspases, are involved in the initiation and execution phases of apoptosis. The activation of the caspases results from the competitive action of the pro-survival and pro-apoptosis Bcl-2-related proteins (Print, C.G. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:12424-12431). A pro-apoptotic signal can activate initiator caspases that trigger a proteolytic caspase cascade, leading to the hydrolysis of target proteins and the classic apoptotic death of the cell. Two active site residues, a cysteine and a histidine, have been

WO 02/097032

PCT/US02/11152

implicated in the catalytic mechanism. Caspases are among the most specific endopeptidases, cleaving after aspartate residues.

Caspases are synthesized as inactive zymogens consisting of one large (p20) and one small (p10) subunit separated by a small spacer region, and a variable N-terminal prodomain. This prodomain interacts with cofactors that can positively or negatively affect apoptosis. An activating signal causes autoproteolytic cleavage of a specific aspartate residue (D297 in the caspase-1 numbering convention) and removal of the spacer and prodomain, leaving a p10/p20 heterodimer. Two of these heterodimers interact via their small subunits to form the catalytically active tetramer. The long proddomains of some caspase family members have been shown to promote dimerization and auto-processing of procaspases. Some caspases contain a "death effector domain" in their prodomain by which they can be recruited into self-activating complexes with other caspases and FADD protein-associated death receptors or the TNF receptor complex. In addition, two dimers from different caspase family members can associate, changing the substrate specificity of the resultant tetramer.

Impaired regulation of apoptosis is associated with loss of neurons in Alzheimer's disease. Alzheimer's disease is a progressive neurodegenerative disorder that is characterized by the formation of senile plaques and neurofibrillary tangles containing amyloid beta peptide. These plaques are found in limbic and association cortices of the brain, including hippocampus, temporal cortices, cingulate cortex, amygdala, nucleus basalis and locus caeruleus. B-amyloid peptide participates in signaling pathways that induce apoptosis and lead to the death of neurons (Kajkowski, C. et al. (2001) J. Biol. Chem. 276:18748-18756). Early in Alzheimer's pathology, physiological changes are visible in the cingulate cortex (Minoshima, S. et al. (1997) Annals of Neurology 42:85-94). In subjects with advanced Alzheimer's disease, accumulating plaques damage the neuronal architecture in limbic areas and eventually cripple the memory process.

Tumor necrosis factor (TNF) and related cytokines induce apoptosis in lymphoid cells. (Reviewed in Nagata, S. (1997) Cell 88:355-365.) Binding of TNF to its receptor triggers a signal transduction pathway that results in the activation of a proteolytic caspase cascade. One such caspase, ICE (Interleukin-1 β converting enzyme), is a cysteine protease comprised of two large and two small subunits generated by ICE auto-cleavage (Dinarello, C. A. (1994) FASEB J. 8:1314-1325). ICE is expressed primarily in monocytes. ICE processes the cytokine precursor, interleukin-1 β , into its active form, which plays a central role in acute and chronic inflammation, bone resorption, myelogenous leukemia, and other pathological processes. ICE and related caspases cause apoptosis when overexpressed in transfected cell lines.

A caspase recruitment domain (CARD) is found within the prodomain of several apical

WO 02/097032

PCT/US02/11152

caspases and is conserved in several apoptosis regulatory molecules such as Apaf-2, RAIDD, and cellular inhibitors of apoptosis proteins (IAPs) (Hofmann, K. et al. (1997) Trends Biochem. Sci. 22:155-157). The regulatory role of CARD in apoptosis may be to allow proteins such as Apaf-1 to associate with caspase-9 (Li, P. et al. (1997) Cell 91:479-489). A human cDNA encoding an
 5 apoptosis repressor with a CARD (ARC) which is expressed in both skeletal and cardiac muscle has been identified and characterized. ARC functions as an inhibitor of apoptosis and interacts selectively with caspases (Koseki, T. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5156-5160). All of these interactions have clear effects on the control of apoptosis (reviewed in Chan S.L. and M.P. Mattson (1999) J. Neurosci. Res. 58:167-190; Salveson, G.S. and V.M. Dixit (1999) Proc. Natl. Acad. Sci.
 10 USA 96:10964-10967).

ES18 was identified as a potential regulator of apoptosis in mouse T-cells (Park, E.J. et al. (1999) Nuc. Acid. Res. 27:1524-1530). ES18 is 428 amino acids in length, contains an N-terminal proline-rich region, an acidic glutamic acid-rich domain, and a putative LXXLL nuclear receptor binding motif. The protein is preferentially expressed in lymph nodes and thymus. The level of ES18
 15 expression increases in T-cell thymoma S49.1 in response to treatment with dexamethasone, staurosporine, or C2-ceramide, which induce apoptosis. ES18 may play a role in stimulating apoptotic cell death in T-cells.

The rat ventral prostate (RVP) is a model system for the study of hormone-regulated apoptosis. RVP epithelial cells undergo apoptosis in response to androgen deprivation. Messenger
 20 RNA (mRNA) transcripts that are up-regulated in the apoptotic RVP have been identified (Briehl, M. M. and Miesfeld, R. L. (1991) Mol. Endocrinol. 5:1381-1388). One such transcript encodes RVP.1, the precise role of which in apoptosis has not been determined. The human homolog of RVP.1, hRVP1, is 89% identical to the rat protein (Katahira, J. et al. (1997) J. Biol. Chem. 272:26652-26658). hRVP1 is 220 amino acids in length and contains four transmembrane domains. hRVP1 is highly
 25 expressed in the lung, intestine, and liver. Interestingly, hRVP1 functions as a low affinity receptor for the *Clostridium perfringens* enterotoxin, a causative agent of diarrhea in humans and other animals.

Cytokine-mediated apoptosis plays an important role in hematopoiesis and the immune response. Myeloid cells, which are the stem cell progenitors of macrophages, neutrophils, erythrocytes, and other blood cells, proliferate in response to specific cytokines such as
 30 granulocyte/macrophage-colony stimulating factor (GM-CSF) and interleukin-3 (IL-3). When deprived of GM-CSF or IL-3, myeloid cells undergo apoptosis. The murine *requiem* (*req*) gene encodes a putative transcription factor required for this apoptotic response in the myeloid cell line FDCP-1 (Gabig, T. G. et al. (1994) J. Biol. Chem. 269:29515-29519). The Req protein is 371 amino

WO 02/097032

PCT/US02/11152

acids in length and contains a nuclear localization signal, a single *Kruppel*-type zinc finger, an acidic domain, and a cluster of four unique zinc-finger motifs enriched in cysteine and histidine residues involved in metal binding. Expression of *req* is not myeloid- or apoptosis-specific, suggesting that additional factors regulate Req activity in myeloid cell apoptosis.

5 Dysregulation of apoptosis has recently been recognized as a significant factor in the pathogenesis of many human diseases. For example, excessive cell survival caused by decreased apoptosis can contribute to disorders related to cell proliferation and the immune response. Such disorders include cancer, autoimmune diseases, viral infections, and inflammation. In contrast, excessive cell death caused by increased apoptosis can lead to degenerative and immunodeficiency disorders such as AIDS, neurodegenerative diseases, and myelodysplastic syndromes. (Thompson, C.B. (1995) *Science* 267:1456-1462.)

Dysregulation of apoptosis has recently been recognized as a significant factor in the pathogenesis of many human diseases. For example, excessive cell survival caused by decreased apoptosis can contribute to disorders related to cell proliferation and the immune response. Such disorders include cancer, autoimmune diseases, viral infections, and inflammation. In contrast, excessive cell death caused by increased apoptosis can lead to degenerative and immunodeficiency disorders such as AIDS, neurodegenerative diseases, and myelodysplastic syndromes. (Thompson, C.B. (1995) *Science* 267:1456-1462.)

Impaired regulation of apoptosis is also associated with loss of neurons in Alzheimer's disease. Alzheimer's disease is a progressive neurodegenerative disorder that is characterized by the formation of senile plaques and neurofibrillary tangles containing amyloid beta peptide. These plaques are found in limbic and association cortices of the brain, including hippocampus, temporal cortices, cingulate cortex, amygdala, nucleus basalis and locus caeruleus. B-amyloid peptide participates in signaling pathways that induce apoptosis and lead to the death of neurons (Kajkowski, C. et al. (2001) *J. Biol. Chem.* 276:18748-18756). Early in Alzheimer's pathology, physiological changes are visible in the cingulate cortex (Minoshima, S. et al. (1997) *Annals of Neurology* 42:85-94). In subjects with advanced Alzheimer's disease, accumulating plaques damage the neuronal architecture in limbic areas and eventually cripple the memory process.

Cancer

30 Cancer remains a major public health concern, and current preventative measures and treatments do not match the needs of most patients. Cancers, also called neoplasias, are characterized by continuous and uncontrolled cell proliferation. They can be divided into three categories: carcinomas, sarcomas, and leukemias. Carcinomas are malignant growths of soft epithelial

WO 02/097032

PCT/US02/11152

cells that may infiltrate surrounding tissues and give rise to metastatic tumors. Sarcomas may be of epithelial origin or arise from connective tissue. Leukemias are progressive malignancies of blood-forming tissue characterized by proliferation of leukocytes and their precursors, and may be classified as myelogenous (granulocyte- or monocyte-derived) or lymphocytic (lymphocyte-derived).

- 5 Tumorigenesis refers to the progression of a tumor's growth from its inception. Malignant cells may be quite similar to normal cells within the tissue of origin or may be undifferentiated (anaplastic). Tumor cells may possess few nuclei or one large polymorphic nucleus. Anaplastic cells may grow in a disorganized mass that is poorly vascularized and as a result contain large areas of ischemic necrosis. Differentiated neoplastic cells may secrete the same proteins as the tissue of origin.
- 10 Cancers grow, infiltrate, invade, and destroy the surrounding tissue through direct seeding of body cavities or surfaces, through lymphatic spread, or through hematogenous spread. Cancer remains a major public health concern and current preventative measures and treatments do not match the needs of most patients. Understanding of the neoplastic process of tumorigenesis can be aided by the identification of molecular markers of prognostic and diagnostic importance.
- 15 Understanding of the neoplastic process can be aided by the identification of molecular markers of prognostic and diagnostic importance. Cancers are associated with oncoproteins which are capable of transforming normal cells into malignant cells. Some oncoproteins are mutant isoforms of the normal protein while others are abnormally expressed with respect to location or level of expression. Normal cell proliferation begins with binding of a growth factor to its receptor on the cell
- 20 membrane, resulting in activation of a signal system that induces and activates nuclear regulatory factors to initiate DNA transcription, subsequently leading to cell division. Classes of oncoproteins known to affect the cell cycle controls include growth factors, growth factor receptors, intracellular signal transducers, nuclear transcription factors, and cell-cycle control proteins. Several types of cancer-specific genetic markers, such as tumor antigens and tumor suppressors, have also been
- 25 identified.

Cancers or malignant tumors, which are characterized by continuous cell proliferation and cell death, can be classified into three categories: carcinomas, sarcomas, and leukemia. Reports show that approximately one in eight women contracts breast cancer and that approximately one in ten men over 50 years of age contracts prostate cancer. (Helzlsouer, K. J. (1994) *Curr. Opin. Oncol.* 6: 541-548;

- 30 Harris, J. R. et al. (1992) *N. Engl. J. Med.* 327:319-328.)

Cancers are associated with the activation of oncogenes which are derived from normal cellular genes. These oncogenes encode oncoproteins which are capable of converting normal cells into malignant cells. Some oncoproteins are mutated isoforms of the normal protein, while other

WO 02/097032

PCT/US02/11152

oncoproteins are abnormally expressed with respect to location or level of expression. The latter category of oncoproteins causes cancer by altering transcriptional control of cell proliferation. Five classes of oncoproteins are known to affect the cell cycle controls. These classes include growth factors, growth factor receptors, intracellular signal transducers, nuclear transcription factors, and cell-cycle control proteins. In some cases, oncogenes can be activated by retroviruses and DNA viruses. Oncogene activation occurs as a consequence of the integration of a viral genome into the DNA of the host cell. In these cases, more than one oncogene, capable of maintaining the infected cell in a condition of continuous cell division, may be activated.

Cancers are characterized by continuous or uncontrolled cell proliferation. Some cancers are associated with suppression of normal apoptotic cell death. Understanding of the neoplastic process can be aided by the identification of molecular markers of prognostic and diagnostic importance. Cancers are associated with oncoproteins which are capable of transforming normal cells into malignant cells. Some oncoproteins are mutant isoforms of the normal protein while others are abnormally expressed with respect to location or level of expression. Normal cell proliferation begins with binding of a growth factor to its receptor on the cell membrane, resulting in activation of a signal system that induces and activates nuclear regulatory factors to initiate DNA transcription, subsequently leading to cell division. Classes of oncoproteins known to affect the cell cycle controls include growth factors, growth factor receptors, intracellular signal transducers, nuclear transcription factors, and cell-cycle control proteins. Several types of cancer-specific genetic markers, such as tumor antigens and tumor suppressors, have also been identified.

Current forms of cancer treatment include the use of immunosuppressive drugs (Morisaki, T., Matsunaga H., et al. (2000) *Anticancer Res.* 20: 3363-3373; Georger, B., Kerr, K., et al. (2001) *Cancer Res.* 61: 1527-1532). The identification of proteins involved in cell signaling, and specifically proteins that act as receptors for immunosuppressant drugs, may facilitate the development of anti-tumor agents. For example, immunophilins are a family of conserved proteins found in both prokaryotes and eukaryotes that bind to immunosuppressive drugs with varying degrees of specificity. One such group of immunophilic proteins is the peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8) family (PPIase, rotamase). These enzymes, first isolated from porcine kidney cortex, accelerate protein folding by catalyzing the cis-trans isomerization of proline imidic peptide bonds in oligopeptides (Fischer, G. and Schmid, F.X. (1990) *Biochemistry* 29: 2205-2212). Included within the immunophilin family are the cyclophilins (e.g., peptidyl-prolyl isomerase A or PPIA) and FK-binding protein (e.g., FKBP) subfamilies. Cyclophilins are multifunctional receptor proteins which participate in signal transduction activities, including those mediated by cyclosporin (or cyclosporine). The PPIase domain

WO 02/097032

PCT/US02/11152

of each family is highly conserved between species. Although structurally distinct, these multifunctional receptor proteins are involved in numerous signal transduction pathways, and have been implicated in folding and trafficking events.

- The immunophilin protein cyclophilin binds to the immunosuppressant drug cyclosporin A.
- 5 FKBP, another immunophilin, binds to FK506 (or rapamycin). Rapamycin is an immunosuppressant agent that arrests cells in the G₁ phase of growth, inducing apoptosis. Like cyclophilin, this macrolide antibiotic (produced by *Streptomyces tsukubaensis*) acts by binding to ubiquitous, predominantly cytosolic immunophilin receptors. These immunophilin/immunosuppressant complexes (e.g., cyclophilin A/cyclosporin A (CypA/CsA) and FKBP12/FK506) achieve their therapeutic results
- 10 through inhibition of the phosphatase calcineurin, a calcium/calmodulin-dependent protein kinase that participates in T-cell activation (Hamilton, G.S. and Steiner, J.P. (1998) *J. Med. Chem.* 41: 5119-5143). The murine fkbp51 gene is abundantly expressed in immunological tissues, including the thymus and T lymphocytes (Baughman, G., Wiederrecht, G.J., et al. (1995) *Molec. Cell. Biol.* 15: 4395-4402). FKBP12/rapamycin-directed immunosuppression occurs through binding to TOR (yeast
- 15 or FRAP (FKBP12-rapamycin-associated protein, in mammalian cells), the kinase target of rapamycin essential for maintaining normal cellular growth patterns. Dysfunctional TOR signaling has been linked to various human disorders including cancer (Metcalf, S.M., Canman, C.E., et al. (1997) *Oncogene* 15: 1635-1642; Emami, S., Le Flock, N., et al. (2001) *FASEB J.* 15: 351-361), and autoimmunity (Damoiseaux, J.G., Beijleveld, L.J., et al. (1996) *Transplantation* 62: 994-1001).
- 20 Several cyclophilin isozymes have been identified, including cyclophilin B, cyclophilin C, mitochondrial matrix cyclophilin, bacterial cytosolic and periplasmic PPIases, and natural-killer cell cyclophilin-related protein possessing a cyclophilin-type PPIase domain, a putative tumor-recognition complex involved in the function of natural killer (NK) cells. These cells participate in the innate cellular immune response by lysing virally-infected cells or transformed cells. NK cells specifically
- 25 target cells that have lost their expression of major histocompatibility complex (MHC) class I genes (common during tumorigenesis), endowing them with the potential for attenuating tumor growth. A 150-kDa molecule has been identified on the surface of human NK cells that possesses a domain which is highly homologous to cyclophilin/peptidyl-prolyl cis-trans isomerase. This cyclophilin-type protein may be a component of a putative tumor-recognition complex, a NK tumor recognition
- 30 sequence (NK-TR) (Anderson, S.K., Gallinger, S., et al. (1993) *Proc. Natl. Acad. Sci. USA* 90: 542-546). The NKTR tumor recognition sequence mediates recognition between tumor cells and large granular lymphocytes (LGLs), a subpopulation of white blood cells (comprised of activated cytotoxic T cells and natural killer cells) capable of destroying tumor targets. The protein product of the NKTR

WO 02/097032

PCT/US02/11152

gene presents on the surface of LGLs and facilitates binding to tumor targets. More recently, a mouse Nkr gene and promoter region have been located on chromosome 9. The gene encodes a NK-cell-specific 150-kDa protein (NK-TR) that is homologous to cyclophilin and other tumor-responsive proteins (Simons-Evelyn, M., Young, H.A. and Anderson, S.K. (1997) Genomics 40: 94-100).

Other proteins that interact with tumorigenic tissue include cytokines such as tumor necrosis factor (TNF). The TNF family of cytokines are produced by lymphocytes and macrophages, and can cause the lysis of transformed (tumor) endothelial cells. Endothelial protein 1 (Edp1) has been identified as a human gene activated transcriptionally by TNF-alpha in endothelial cells, and a TNF-alpha inducible Edp1 gene has been identified in the mouse (Swift, S., Blackburn, C., et al. (1998) Biochim. Biophys. Acta 1442: 394-398).

Oncogenes

Many oncogenes have been identified and characterized. These include growth factors such as *sis*, receptors such as *erbA*, *erbB*, *neu*, and *ros*, intracellular receptors such as *src*, *yes*, *fps*, *abl*, and *met*, protein-serine/threonine kinases such as *mos* and *raf*, nuclear transcription factors such as *jun*, *fos*, *myc*, *N-myc*, *myb*, *ski*, and *rel*, cell cycle control proteins such as *RB* and *p53*, mutated tumor-suppressor genes such as *mdm2*, *Cip1*, *p16*, and *cyclin D*, *ras*, *set*, *can*, *sec*, and *gag R10*. In particular, FOS encoded by *fos*, is a leucine-zipper-containing phosphoprotein located in the nucleus of cells. FOS forms a non-covalent complex with several other proteins to activate the transcription of growth-promoting proteins. (Bohmann, D. et al. (1987) Science 238:1386-1392; Cohen, D.R. and Curran, T. (1988) Mol. Cell. Biol. 8: 2063-2069; and van Straaten, F. et al. (1983) Proc. Natl. Acad. Sci. 80: 3188-3187.) *can* is a putative human oncogene associated with myeloid leukemogenesis and is activated as an oncogene by fusion of its 3' half with other genes such as *set*. (von Lindern, M. et al. (1992) Mol. Cell. Biol. 12: 3346-3355.) SET, encoded by *set*, is shown to be a potent inhibitor of phosphatase 2A, a serine/threonine phosphatase that regulates diverse cellular processes. (Li, M. et al. (1996) J. Biol. Chem. 271: 11059-11062.) The *Xenopus* homolog of SET, NAP1, is found to interact specifically with B-type cyclins and plays an essential role in cell cycle regulation. (Kellogg, D. R. et al. (1995) J. Cell Biol. 130: 661-673.) SEC is the gene product of *sec* and is an oncoprotein active in tumors of secretory epithelium. (Lane, M.A. et al. (1990) Nuc. Acids Res. 18: 3068.) *gag R10* is a leucine zipper-containing cytoplasmic protein of 23 kDa identified from chicken embryonic neuroretina cells and is encoded by a chimeric mRNA, RAV-1, which is capable of inducing cells to continuous cell proliferation. (Proux, V. et al. (1996) J. Biol. Chem. 271: 30790-30797.) S-100 are a family of small dimeric acidic calcium and zinc-binding proteins expressed abundantly in brain. These

WO 02/097032

PCT/US02/11152

proteins play important roles in cell growth and differentiation, cell cycle regulation, and metabolic control. (Moncrief, N.D. et al. (1990) J. Mol. Evol. 30: 522-562; and Wicki, R. et al. (1996) Biochem. Biophys. Res. Commun. 227: 594-599.) rad1 is a yeast protein involved in DNA repair and recombination. (Sunnerhagen, P. et al. (1990) Mol. Cell. Biol. 10: 3750-3760.) Alpha-L-fucosidase is a lysosomal enzyme which hydrolyzes alpha-1,6 bond between fucose and the N-acetylglucosamine of the carbohydrate moieties of glycoproteins. Deficiency of alpha-L-fucosidase results in fucosidosis, a lysosomal storage disease. (Herissat, B. (1991) Biochem. J. 280: 309-316.)

Oncoproteins are encoded by genes, called oncogenes, that are derived from genes that normally control cell growth and development. Many oncogenes have been identified and characterized. These include growth factors such as *sis*, receptors such as *erbA*, *erbB*, *neu*, and *ros*, intracellular receptors such as *src*, *yes*, *fps*, *abl*, and *met*, protein-serine/threonine kinases such as *mos* and *raf*, nuclear transcription factors such as *jun*, *fos*, *myc*, *N-myc*, *myb*, *ski*, and *rel*, cell cycle control proteins such as *RB* and *p53*, mutated tumor-suppressor genes such as *mdm2*, *Cip1*, *p16*, and *cyclin D*, *ras*, *set*, *can*, *sec*, and *gag R10*.

Viral oncogenes are integrated into the human genome after infection of human cells by certain viruses. Examples of viral oncogenes include v-src, v-abl, and v-fps. Transformation of normal genes to oncogenes may also occur by chromosomal translocation. The Philadelphia chromosome, characteristic of chronic myeloid leukemia and a subset of acute lymphoblastic leukemias, results from a reciprocal translocation between chromosomes 9 and 22 that moves a truncated portion of the proto-oncogene c-abl to the breakpoint cluster region (bcr) on chromosome 22. The hybrid c-abl-bcr gene encodes a chimeric protein that has tyrosine kinase activity. In chronic myeloid leukemia, the chimeric protein has a molecular weight of 210 kd, whereas in acute leukemias a more active 180 kd tyrosine kinase is formed (Robbins, S.L. et al. (1994) Pathologic Basis of Disease, W.B. Saunders Co., Philadelphia PA).

The Wnt gene family of secreted signaling molecules is highly conserved throughout eukaryotic cells. Members of the Wnt family are involved in regulating chondrocyte differentiation within the cartilage template. Wnt-5a, Wnt-5b and Wnt-4 genes are expressed in chondrogenic regions of the chicken limb, Wnt-5a being expressed in the perichondrium (mesenchymal cells immediately surrounding the early cartilage template). Wnt-5a misexpression delays the maturation of chondrocytes and the onset of bone collar formation in chicken limb (Hartmann, C. and Tabin, C.J. (2000) *Development* 127:3141-3159).

RRP22protein/RAS-related proteins

Signal transduction is the general process by which cells respond to extracellular signals. In

WO 02/097032

PCT/US02/11152

typical signal transduction pathways, binding of a signaling molecule such as a hormone, neurotransmitter, or growth factor to a cell membrane receptor is coupled to the action of an intracellular second messenger. G protein-coupled receptors (GPCRs) control intracellular processes through the activation of guanine nucleotide-binding proteins (G proteins). G proteins are

5 heterotrimeric and consist of α , β , and γ subunits. The α subunits contain a guanine nucleotide binding domain and have GTPase activity. When GTP binds to α subunits, it dissociates from the β and γ subunits and interacts with cellular target molecules. Hydrolysis of GTP to GDP serves as a molecular switch controlling the interactions of the subunit with other proteins. The GDP bound form of the α subunit dissociates from its cellular target and reassociates with the β and γ subunits. A

10 number of accessory proteins modulate G protein function by controlling their nucleotide phosphorylation state or membrane association. These regulatory molecules include exchange factors (GEFs) which stimulate GDP-GTP exchange, GTPase activating proteins (GAPs) which promote GTP hydrolysis, and guanine nucleotide dissociation inhibitors (GDIs) which inhibit guanine nucleotide dissociation and stabilize the GDP-bound form. G proteins can be classified into at least five

15 subfamilies: Ras, Rho, Ran, Rab, and ADP-ribosylation factor, and they regulate various cell functions including cell growth and differentiation, cytoskeletal organization, and intracellular vesicle transport and secretion.

The Ras superfamily of small GTPases is involved in the regulation of a wide range of cellular signaling pathways. Ras family proteins are membrane-associated proteins acting as molecular

20 switches that bind GTP and GDP, hydrolyzing GTP to GDP. In the active GTP-bound state Ras family proteins interact with a variety of cellular targets to activate downstream signaling pathways. For example, members of the Ras subfamily are essential in transducing signals from receptor tyrosine kinases (RTKs) to a series of serine/threonine kinases which control cell growth and differentiation. Activated Ras genes were initially found in human cancers and subsequent studies confirmed that Ras

25 function is critical in the determination of whether cells continue to grow or become terminally differentiated (Barbacid, M. (1987) *Annu. Rev. Biochem.* 56:779-827, Treisman, R. (1994) *Curr. Opin. Genet. Dev.* 4:96-98). Mutant Ras proteins, which bind but can not hydrolyze GTP, are permanently activated, and cause continuous cell proliferation or cancer.

The Ras subfamily transduces signals from tyrosine kinase receptors, non-tyrosine kinase

30 receptors, and heterotrimeric GPCRs (Fantl, W. J. et al. (1993) *Annu. Rev. Biochem.* 62:453-481; Woodrow, M. A. et al. (1993) *J. Immunol.* 150:3853-3861; and van Corven, E. J. et al. (1993) *Proc. Natl. Acad. Sci.* 90:1257-1261). Stimulation of cell surface receptors activates Ras which, in turn, activates cytoplasmic kinases that control cell growth and differentiation. The first Ras targets

WO 02/097032

PCT/US02/11152

identified were the Raf kinases (Avruch, J. et al. (1994) Trends Biochem. Sci. 19:279-283).

Interaction of Ras and Raf leads to activation of the MAP kinase cascade of serine/threonine kinases, which activate key transcription factors that control gene expression and protein synthesis (Barbacid, M. (1987) Ann. Rev. Biochem. 56:779-827; Treisman, R. (1994) Curr. Opin. Genet. Dev. 4:96-101).

- 5 Mutated Ras proteins, which bind but do not hydrolyze GTP, are constitutively activated, and cause continuous cell proliferation and cancer (Bos, J. L. (1989) Cancer Res. 49:4682-4689; Grunicke, H. H. and Maly, K. (1993) Crit. Rev. Oncog. 4:389-402).

Many oncogenes have been identified and characterized. These include growth factors such as *sis*, receptors such as *erbA*, *erbB*, *neu*, and *ros*, intracellular receptors such as *src*, *yes*, *fps*, *abl*,
10 and *met*, protein-serine/threonine kinases such as *mos* and *raf*, nuclear transcription factors such as *jun*, *fos*, *myc*, *N-myc*, *myb*, *ski*, and *rel*, cell cycle control proteins such as *RB* and *p53*, mutated tumor-suppressor genes such as *mdm2*, *Cip1*, *p16*, and *cyclin D*, *ras*, *set*, *can*, *sec*, and *gag* *R10*.

In particular, FOS encoded by *fos*, is a leucine-zipper-containing phosphoprotein located in the nucleus of cells. FOS forms a non-covalent complex with several other proteins to activate the transcription of
15 growth-promoting proteins. (Bohmann, D. et al. (1987) Science 238:1386-1392; Cohen, D.R. and Curran, T. (1988) Mol. Cell. Biol. 8: 2063-2069; and van Straaten, F. et al. (1983) Proc. Natl. Acad. Sci. 80: 3188-3187.) *can* is a putative human oncogene associated with myeloid leukemogenesis and is activated as an oncogene by fusion of its 3' half with other genes such as *set*. (von Lindern, M. et

- al. (1992) Mol. Cell. Biol. 12: 3346-3355.) SET, encoded by *set*, is shown to be a potent inhibitor of
20 phosphatase 2A, a serine/threonine phosphatase that regulates diverse cellular processes. (Li, M. et al. (1996) J. Biol. Chem. 271: 11059-11062.) The *Xenopus* homolog of SET, NAPI, is found to interact specifically with B-type cyclins and plays an essential role in cell cycle regulation. (Kellogg, D. R. et al. (1995) J. Cell Biol. 130: 661-673.) SEC is the gene product of *sec* and is an oncoprotein active in tumors of secretory epithelium. (Lane, M.A. et al. (1990) Nuc. Acids Res. 18: 3068.) *gag*
25 *R10* is a leucine zipper-containing cytoplasmic protein of 23 kDa identified from chicken embryonic neuroretina cells and is encoded by a chimeric mRNA, RAV-1, which is capable of inducing cells to continuous cell proliferation. (Proux, V. et al. (1996) J. Biol. Chem. 271: 30790-30797.) S-100 are a family of small dimeric acidic calcium and zinc-binding proteins expressed abundantly in brain. These proteins play important roles in cell growth and differentiation, cell cycle regulation, and metabolic
30 control. (Moncrief, N.D. et al. (1990) J. Mol. Evol. 30: 522-562; and Wicki, R. et al. (1996) Biochem. Biophys. Res. Commun. 227: 594-599.) *rad1* is a yeast protein involved in DNA repair and recombination. (Sunnerhagen, P. et al. (1990) Mol. Cell. Biol. 10: 3750-3760.) Alpha-L-fucosidase is a lysosomal enzyme which hydrolyzes alpha-1,6 bond between fucose and the N-acetylglucosamine of

WO 02/097032

PCT/US02/11152

the carbohydrate moieties of glycoproteins. Deficiency of alpha-L-fucosidase results in fucosidosis, a lysosomal storage disease. (Herissat, B. (1991) *Biochem. J.* 280: 309-316.)

Ras regulates other signaling pathways by direct interaction with different cellular targets (Katz, M. E. and McCormick, F. (1997) *Curr. Opin. Genet. Dev.* 7:75-79). One such target is RalGDS, a guanine nucleotide dissociation stimulator for the Ras-like GTPase, Ral (Albright, C. F. et al. (1993) *EMBO J.* 12:339-347). RalGDS couples the Ras and Ral signaling pathways. Epidermal growth factor (EGF) stimulates the association of RalGDS with Ras in mammalian cells, which activates the GEF activity of RalGDS (Kikuchi, A. and Williams, L. T. (1996) *J. Biol. Chem.* 271:588-594; Urano, T. et al. (1996) *EMBO J.* 15:810-816). Ral activation by Ral-GDS leads to activation of Src, a tyrosine kinase that phosphorylates other molecules including transcription factors and components of the actin cytoskeleton (Goi, T. et al. (2000) *EMBO J.* 19:623-630). Ral interacts with a number of signaling molecules including Ral-binding protein, a GAP for the Rho-like GTPases; Cdc42 and Rac, which regulate cytoskeletal rearrangement; and phospholipase D1, which is involved in vesicular trafficking (Feig, L. A. et al. (1996) *Trends Biochem. Sci.* 21:438-441; Voss, M. et al. (1999) *J. Biol. Chem.* 274:34691-34698).

Nore1 was identified from a yeast two-hybrid screen as a protein that interacts with Ras and Ras-related protein, Rap1b (Vavvas, D. et al. (1998) *J. Biol. Chem.* 273:5439-5442). It is a basic protein (pI=9.4) of 413 amino acids that contains a cysteine-histidine-rich region predicted to be a diacylglycerol/phorbol ester binding site, a proline-rich region at its N-terminus that may be an SH3 binding domain, and a Ras/Rap binding domain located at its C-terminus. Nore1 binds Ras *in vitro* in a GTP-dependent manner. Experiments *in vivo* show that the association of Nore1 with Ras is dependent on EGF and 12-*O*-tetradecanoylphorbol-13-acetate activation in COS-7 cells and on EGF in KB cells.

Ras and other G proteins play roles in regulating the immune inflammatory response. Granulocytes, which include basophils, eosinophils, and neutrophils, play critical roles in inflammation. Eosinophils release toxic granule proteins, which kill microorganisms, and secrete prostaglandins, leukotrienes and cytokines, which amplify the inflammatory response. They sustain inflammation in allergic reactions and their malfunction can cause asthma and other allergic diseases. Interleukin-5 is a cytokine that regulates the growth, activation, and survival of eosinophils. The signal transduction mechanism of IL-5 in eosinophils involves the Ras-MAP kinase and Jak-Stat pathways (Pazdrak, K. et al. (1995) *J. Exp. Med.* 181:1827-1834; Adachi, T. and Ala, R. (1998) *Am. J. Physiol.* 275:C623-633). Raf-1 kinase activation by Ras is implicated in eosinophil degranulation.

Neutrophils migrate to inflammatory sites where they eliminate pathogens by phagocytosis and

WO 02/097032

PCT/US02/11152

- release toxic products from their granules that kill microorganisms. G proteins, including Ras, Ral, Rac1, and Rap1 regulate neutrophil function (M'Rabet, L. et al. (1999) J. Biol. Chem. 274:21847-21852). Rac1 may be involved in the respiratory burst of neutrophils. Ras and Rap1 are activated in response to the chemotactic agent, formyl methionine leucine phenylalanine (fMLP); the lipid mediator, platelet activating factor (PAF); and the cytokine, granulocyte-macrophage colony-stimulating factor (GM-CSF). Both Ras and Rap1 appear to play roles in neutrophil activation. Ral is activated by fMLP and PAF but not by GM-CSF and may be involved in chemotaxis, phagocytosis, or degranulation. Impairment of neutrophil function is associated with various inflammatory and autoimmune diseases.
- RRP22 defines a new subgroup whose expression is limited to the central nervous system. The genes are located in the CpG-rich q12 region of chromosome 22 within a 40-kb region bounded by the EWS and BAM22 genes (Zucman-Rossi, J. et al. (1996) Genomics 38:247-254).
- Activation of Ras family proteins is catalyzed by guanine nucleotide exchange factors (GEFs) which catalyze the dissociation of bound GDP and subsequent binding of GTP. A recently discovered RalGEF-like protein, RGL3, interacts with both Ras and the related protein Rit. Constitutively active Rit, like Ras, can induce oncogenic transformation, although since Rit fails to interact with most known Ras effector proteins, novel cellular targets may be involved in Rit transforming activity. RGL3 interacts with both Ras and Rit, and thus may act as a downstream effector for these proteins (Shao, H. and Andres, D.A. (2000) J. Biol. Chem. 275:26914-26924).
- Tumor antigens
- Tumor antigens are cell surface molecules that are differentially expressed in tumor cells relative to non-tumor tissues. Tumor antigens make tumor cells immunologically distinct from normal cells and are potential diagnostics for human cancers. Several monoclonal antibodies have been identified which react specifically with cancerous cells such as T-cell acute lymphoblastic leukemia and neuroblastoma (Minegishi et al. (1989) Leukemia Res. 13:43-51; Takagi et al. (1995) Int. J. Cancer 61:706-715). In addition, the discovery of high level expression of the HER2 gene in breast tumors has led to the development of therapeutic treatments (Liu et al. (1992) Oncogene 7: 1027-1032; Kern (1993) Am. J. Respir. Cell Mol. Biol. 9:448-454). Tumor antigens are found on the cell surface and have been characterized either as membrane proteins or glycoproteins. For example, MAGE genes encode a family of tumor antigens recognized on melanoma cell surfaces by autologous cytolytic T lymphocytes. Among the 12 human MAGE genes isolated, half are differentially expressed in tumors of various histological types (De Plaen et al. (1994) Immunogenetics 40:360-369). None of the 12 MAGE genes, however, is expressed in healthy tissues except testis and placenta.

WO 02/097032

PCT/US02/11152

Breast Cancer

There are more than 180,000 new cases of breast cancer diagnosed each year, and the mortality rate for breast cancer approaches 10% of all deaths in females between the ages of 45-54 (K. Gish (1999) AWIS Magazine 28:7-10). However the survival rate based on early diagnosis of localized breast cancer is extremely high (97%), compared with the advanced stage of the disease in which the tumor has spread beyond the breast (22%). Current procedures for clinical breast examination are lacking in sensitivity and specificity, and efforts are underway to develop comprehensive gene expression profiles for breast cancer that may be used in conjunction with conventional screening methods to improve diagnosis and prognosis of this disease (Perou, C.M. et al. (2000) Nature 406:747-752).

Mutations in two genes, BRCA1 and BRCA2, are known to greatly predispose a woman to breast cancer and may be passed on from parents to children (Gish, K. (1999) AWIS Magazine 28:7-10). However, this type of hereditary breast cancer accounts for only about 5% to 9% of breast cancers, while the vast majority of breast cancer is due to non-inherited mutations that occur in breast epithelial cells.

The relationship between expression of epidermal growth factor (EGF) and its receptor, EGFR, to human mammary carcinoma has been particularly well studied. (See Khazaie, K. et al. (1993) Cancer and Metastasis Rev. 12:255-274, and references cited therein for a review of this area.) Overexpression of EGFR, particularly coupled with down-regulation of the estrogen receptor, is a marker of poor prognosis in breast cancer patients. In addition, EGFR expression in breast tumor metastases is frequently elevated relative to the primary tumor, suggesting that EGFR is involved in tumor progression and metastasis. This is supported by accumulating evidence that EGF has effects on cell functions related to metastatic potential, such as cell motility, chemotaxis, secretion and differentiation. Changes in expression of other members of the erbB receptor family, of which EGFR is one, have also been implicated in breast cancer. The abundance of erbB receptors, such as HER-2/neu, HER-3, and HER-4, and their ligands in breast cancer points to their functional importance in the pathogenesis of the disease, and may therefore provide targets for therapy of the disease (Bacus, S. S. et al. (1994) Am. J. Clin. Pathol. 102:S13-S24). Other known markers of breast cancer include a human secreted frizzled protein mRNA that is downregulated in breast tumors; the matrix G1a protein which is overexpressed in human breast carcinoma cells; Drg1 or RTP, a gene whose expression is diminished in colon, breast, and prostate tumors; maspin, a tumor suppressor gene downregulated in invasive breast carcinomas; and CaN19, a member of the S100 protein family, all of which are down regulated in mammary carcinoma cells relative to normal mammary epithelial cells

WO 02/097032

PCT/US02/11152

(Zhou, Z. et al. (1998) *Int. J. Cancer* 78:95-99; Chen, L. et al. (1990) *Oncogene* 5:1391-1395; Ulrix, W. et al. (1999) *FEBS Lett* 455:23-26; Sager, R. et al. (1996) *Curr. Top. Microbiol. Immunol.* 213:51-64; and Lee, S. W. et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:2504-2508).

Cell lines derived from human mammary epithelial cells at various stages of breast cancer

- 5 provide a useful model to study the process of malignant transformation and tumor progression as it has been shown that these cell lines retain many of the properties of their parental tumors for lengthy culture periods (Wistuba, I.I. et al. (1998) *Clin. Cancer Res.* 4:2931-2938). Such a model is particularly useful for comparing phenotypic and molecular characteristics of human mammary epithelial cells at various stages of malignant transformation.

10 Tumor suppressors

Tumor suppressor genes are generally defined as genetic elements whose loss or inactivation contributes to the deregulation of cell proliferation and the pathogenesis and progression of cancer.

Tumor suppressor genes normally function to control or inhibit cell growth in response to stress and to limit the proliferative life span of the cell. Several tumor suppressor genes have been identified

- 15 including the genes encoding the retinoblastoma (Rb) protein, p53, and the breast cancer 1 and 2 proteins (BRCA1 and BRCA2). Mutations in these genes are associated with acquired and inherited genetic predisposition to the development of certain cancers.

The role of p53 in the pathogenesis of cancer has been extensively studied. (Reviewed in Aggarwal, M. L. et al. (1998) *J. Biol. Chem.* 273:1-4; Levine, A. (1997) *Cell* 88:323-331.) About 50%

- 20 of all human cancers contain mutations in the p53 gene. These mutations result in either the absence of functional p53 or, more commonly, a defective form of p53 which is overexpressed. p53 is a transcription factor that contains a central core domain required for DNA binding. Most cancer-associated mutations in p53 localize to this domain. In normal proliferating cells, p53 is expressed at low levels and is rapidly degraded. p53 expression and activity is induced in response to DNA
25 damage, abortive mitosis, and other stressful stimuli. In these instances, p53 induces apoptosis or arrests cell growth until the stress is removed. Downstream effectors of p53 activity include apoptosis-specific proteins and cell cycle regulatory proteins, including Rb, oncogene products, cyclins, and cell cycle-dependent kinases.

- The metastasis-suppressor gene KAI1 (CD82) has been reported to be related to the tumor
30 suppressor gene p53. KAI1 is involved in the progression of human prostatic cancer and possibly lung and breast cancers when expression is decreased. KAI1 encodes a member of a structurally distinct family of leukocyte surface glycoproteins. The family is known as either the tetraspan transmembrane protein family or transmembrane 4 superfamily (TM4SF) as the members of this

WO 02/097032

PCT/US02/11152

family span the plasma membrane four times. The family is composed of integral membrane proteins having a N-terminal membrane-anchoring domain which functions as both a membrane anchor and a translocation signal during protein biosynthesis. The N-terminal membrane-anchoring domain is not cleaved during biosynthesis. TM4SF proteins have three additional transmembrane regions, seven or
 5 more conserved cysteine residues, are similar in size (218 to 284 residues), and all have a large extracellular hydrophilic domain with three potential N-glycosylation sites. The promoter region contains many putative binding motifs for various transcription factors, including five AP2 sites and nine Sp1 sites. Gene structure comparisons of KAI1 and seven other members of the TM4SF indicate that the splicing sites relative to the different structural domains of the predicted proteins are
 10 conserved. This suggests that these genes are related evolutionarily and arose through gene duplication and divergent evolution (Levy, S. et al. (1991) *J. Biol. Chem.* 266:14597-14602; Dong, J.T. et al. (1995) *Science* 268:884-886; Dong, J.T. et al., (1997) *Genomics* 41:25-32).

The Leucine-rich gene-Glioma Inactivated (LGI1) protein shares homology with a number of transmembrane and extracellular proteins which function as receptors and adhesion proteins. LGI1 is
 15 encoded by an LLR (leucine-rich, repeat-containing) gene and maps to 10q24. LGI1 has four LLRs which are flanked by cysteine-rich regions and one transmembrane domain (Somerville, R.P., et al. (2000) *Mamm. Genome* 11:622-627). LGI1 expression is seen predominantly in neural tissues, especially brain. The loss of tumor suppressor activity is seen in the inactivation of the LGI1 protein which occurs during the transition from low to high-grade tumors in malignant gliomas. The reduction
 20 of LGI1 expression in low grade brain tumors and its significant reduction or absence of expression in malignant gliomas suggests that it could be used for diagnosis of glial tumor progression (Chernova, O.B., et al. (1998) *Oncogene* 17:2873-2881).

The ST13 tumor suppressor was identified in a screen for factors related to colorectal carcinomas by subtractive hybridization between cDNA of normal mucosal tissues and mRNA of
 25 colorectal carcinoma tissues (Cao, J. et al. (1997) *J. Cancer Res. Clin. Oncol.* 123:447-451). ST13 is down-regulated in human colorectal carcinomas.

Mutations in the von Hippel-Lindau (VHL) tumor suppressor gene are associated with retinal and central nervous system hemangioblastomas, clear cell renal carcinomas, and pheochromocytomas (Hoffman, M. et al. (2001) *Hum. Mol. Genet.* 10:1019-1027; Kamada, M. (2001) *Cancer Res.*
 30 61:4184-4189). Tumor progression is linked to defects or inactivation of the VHL gene. VHL regulates the expression of transforming growth factor- α , the GLUT-1 glucose transporter and vascular endothelial growth factor. The VHL protein associates with elongin B, elongin C, Cul2 and Rbx1 to form a complex that regulates the transcriptional activator hypoxia-inducible factor (HIF).

WO 02/097032

PCT/US02/11152

HIF induces genes involved in angiogenesis such as vascular endothelial growth factor and platelet-derived growth factor B. Loss of control of HIF caused by defects in VHL results in the excessive production of angiogenic peptides. VHL may play roles in inhibition of angiogenesis, cell cycle control, fibronectin matrix assembly, cell adhesion, and proteolysis.

5 Mutations in tumor suppressor genes are a common feature of many cancers and often appear to affect a critical step in the pathogenesis and progression of tumors. Accordingly, Chang, F. et al. (1995; J. Clin. Oncol. 13: 1009-1022) suggest that it may be possible to use either the gene or an antibody to the expressed protein 1) to screen patients at increased risk for cancer, 2) to aid in diagnosis made by traditional methods, and 3) to assess the prognosis of individual cancer patients. In addition, Hamada, K et al. (1996; Cancer Res. 56:3047-3054) are investigating the introduction of p53 into cervical cancer cells via an adenoviral vector as an experimental therapy for cervical cancer.

The PR-domain genes were recently recognized as playing a role in human tumorigenesis. PR-domain genes normally produce two protein products: the PR-plus product, which contains the PR domain, and the PR-minus product which lacks this domain. In cancer cells, PR-plus is disrupted or overexpressed, while PR-minus is present or overexpressed. The imbalance in the amount of these two proteins appears to be an important cause of malignancy (Jiang, G.L. and Huang, S. (2000) Histol. Histopathol. 15:109-117).

Many neoplastic disorders in humans can be attributed to inappropriate gene transcription. Malignant cell growth may result from either excessive expression of tumor promoting genes or insufficient expression of tumor suppressor genes (Cleary, M.L. (1992) Cancer Surv. 15:89-104). Chromosomal translocations may also produce chimeric loci which fuse the coding sequence of one gene with the regulatory regions of a second unrelated gene. An important class of transcriptional regulators are the zinc finger proteins. The zinc finger motif, which binds zinc ions, generally contains tandem repeats of about 30 amino acids consisting of periodically spaced cysteine and histidine residues. Examples of this sequence pattern include the C2H2-type, C4-type, and C3HC4-type zinc fingers, and the PHD domain (Lewin, supra; Aasland, R., et al. (1995) Trends Biochem. Sci. 20:56-59). One clinically relevant zinc-finger protein is WT1, a tumor-suppressor protein that is inactivated in children with Wilm's tumor. The oncogene bcl-6, which plays an important role in large-cell lymphoma, is also a zinc-finger protein (Papavassiliou, A. G. (1995) N. Engl. J. Med. 332:45-47).

30 Tumor responsive proteins

Cancers, also called neoplasias, are characterized by continuous and uncontrolled cell proliferation. They can be divided into three categories: carcinomas, sarcomas, and leukemias. Carcinomas are malignant growths of soft epithelial cells that may infiltrate surrounding tissues and

WO 02/097032

PCT/US02/11152

give rise to metastatic tumors. Sarcomas may be of epithelial origin or arise from connective tissue. Leukemias are progressive malignancies of blood-forming tissue characterized by proliferation of leukocytes and their precursors, and may be classified as myelogenous (granulocyte- or monocyte-derived) or lymphocytic (lymphocyte-derived). Tumorigenesis refers to the progression of a tumor's growth from its inception. Malignant cells may be quite similar to normal cells within the tissue of origin or may be undifferentiated (anaplastic). Tumor cells may possess few nuclei or one large polymorphic nucleus. Anaplastic cells may grow in a disorganized mass that is poorly vascularized and as a result contains large areas of ischemic necrosis. Differentiated neoplastic cells may secrete the same proteins as the tissue of origin. Cancers grow, infiltrate, invade, and destroy the surrounding tissue through direct seeding of body cavities or surfaces, through lymphatic spread, or through hematogenous spread. Cancer remains a major public health concern and current preventative measures and treatments do not match the needs of most patients. Understanding of the neoplastic process of tumorigenesis can be aided by the identification of molecular markers of prognostic and diagnostic importance.

Current forms of cancer treatment include the use of immunosuppressive drugs (Morisaki, T. et al. (2000) *Anticancer Res.* 20: 3363-3373; Georger, B. et al. (2001) *Cancer Res.* 61: 1527-1532). The identification of proteins involved in cell signaling, and specifically proteins that act as receptors for immunosuppressant drugs, may facilitate the development of anti-tumor agents. For example, immunophilins are a family of conserved proteins found in both prokaryotes and eukaryotes that bind to immunosuppressive drugs with varying degrees of specificity. One such group of immunophilic proteins is the peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8) family (PPIase, rotamase). These enzymes, first isolated from porcine kidney cortex, accelerate protein folding by catalyzing the cis-trans isomerization of proline imidic peptide bonds in oligopeptides (Fischer, G. and Schmid, F.X. (1990) *Biochemistry* 29: 2205-2212). Included within the immunophilin family are the cyclophilins (e.g., peptidyl-prolyl isomerase A or PPLA) and FK-binding protein (e.g., FKBP) subfamilies. Cyclophilins are multifunctional receptor proteins which participate in signal transduction activities, including those mediated by cyclosporin (or cyclosporine). The PPIase domain of each family is highly conserved between species. Although structurally distinct, these multifunctional receptor proteins are involved in numerous signal transduction pathways, and have been implicated in folding and trafficking events.

The immunophilin protein cyclophilin binds to the immunosuppressant drug cyclosporin A. FKBP, another immunophilin, binds to FK506 (or rapamycin). Rapamycin is an immunosuppressant agent that arrests cells in the G₁ phase of growth, inducing apoptosis. Like cyclophilin, this macrolide

WO 02/097032

PCT/US02/11152

antibiotic (produced by *Streptomyces tsukubaensis*) acts by binding to ubiquitous, predominantly cytosolic immunophilin receptors. These immunophilin/immunosuppressant complexes (e.g., cyclophilin A/cyclosporin A (CypA/CsA) and FKBP12/FK506) achieve their therapeutic results through inhibition of the phosphatase calcineurin, a calcium/calmodulin-dependent protein kinase that participates in T-cell activation (Hamilton, G.S. and Steiner, J.P. (1998) J. Med. Chem. 41: 5119-5143). The murine fkbp51 gene is abundantly expressed in immunological tissues, including the thymus and T lymphocytes (Baughman, G. et al. (1995) Molec. Cell. Biol. 15: 4395-4402). FKBP12/rapamycin-directed immunosuppression occurs through binding to TOR (yeast) or FRAP (FKBP12-rapamycin-associated protein, in mammalian cells), the kinase target of rapamycin essential for maintaining normal cellular growth patterns. Dysfunctional TOR signaling has been linked to various human disorders including cancer (Metcalf, S.M. et al. (1997) Oncogene 15: 1635-1642; Emami, S. et al. (2001) FASEB J. 15: 351-361), and autoimmunity (Damoiseaux, J.G. et al. (1996) Transplantation 62: 994-1001).

Several cyclophilin isozymes have been identified, including cyclophilin B, cyclophilin C, mitochondrial matrix cyclophilin, bacterial cytosolic and periplasmic PPIases, and natural-killer cell cyclophilin-related protein possessing a cyclophilin-type PPIase domain, a putative tumor-recognition complex involved in the function of natural killer (NK) cells. These cells participate in the innate cellular immune response by lysing virally-infected cells or transformed cells. NK cells specifically target cells that have lost their expression of major histocompatibility complex (MHC) class I genes (common during tumorigenesis), endowing them with the potential for attenuating tumor growth. A 150-kDa molecule has been identified on the surface of human NK cells that possesses a domain which is highly homologous to cyclophilin/peptidyl-prolyl cis-trans isomerase. This cyclophilin-type protein may be a component of a putative tumor-recognition complex, a NK tumor recognition sequence (NK-TR) (Anderson, S.K. et al. (1993) Proc. Natl. Acad. Sci. USA 90: 542-546). The NKTR tumor recognition sequence mediates recognition between tumor cells and large granular lymphocytes (LGLs), a subpopulation of white blood cells (comprised of activated cytotoxic T cells and natural killer cells) capable of destroying tumor targets. The protein product of the NKTR gene presents on the surface of LGLs and facilitates binding to tumor targets. More recently, a mouse Nktr gene and promoter region have been located on chromosome 9. The gene encodes a NK-cell-specific 150-kDa protein (NK-TR) that is homologous to cyclophilin and other tumor-responsive proteins (Simons-Evelyn, M. et al. (1997) Genomics 40: 94-100).

Other proteins that interact with tumorigenic tissue include cytokines such as tumor necrosis factor (TNF). The TNF family of cytokines are produced by lymphocytes and macrophages, and can

WO 02/097032

PCT/US02/11152

cause the lysis of transformed (tumor) endothelial cells. Endothelial protein 1 (Edp1) has been identified as a human gene activated transcriptionally by TNF-alpha in endothelial cells, and a TNF-alpha inducible Edp1 gene has been identified in the mouse (Swift, S. et al. (1998) Biochim. Biophys. Acta 1442: 394-398).

5 Aging and Senescence

Studies of the aging process or senescence have shown a number of characteristic cellular and molecular changes (Fauci et al. (1998) Harrison's Principles of Internal Medicine, McGraw-Hill, New York NY, p.37). These characteristics include increases in chromosome structural abnormalities, DNA cross-linking, incidence of single-stranded breaks in DNA, losses in DNA methylation, and degradation of telomere regions. In addition to these DNA changes, post-translational alterations of proteins increase including, deamidation, oxidation, cross-linking, and nonenzymatic glycation. Still further molecular changes occur in the mitochondria of aging cells through deterioration of structure. These changes eventually contribute to decreased function in every organ of the body.

15 Lung Cancer

Lung cancer is the leading cause of cancer death for men and the second leading cause of cancer death for women in the U.S. Lung cancers are divided into four histopathologically distinct groups. Three groups (squamous cell carcinoma, adenocarcinoma, and large cell carcinoma) are classified as non-small cell lung cancers (NSCLCs). The fourth group of cancers is referred to as small cell lung cancer (SCLC). Deletions on chromosome 3 are common in this disease and are thought to indicate the presence of a tumor suppressor gene in this region. Activating mutations in K-ras are commonly found in lung cancer and are the basis of one of the mouse models for the disease.

Steroid Hormones

Glucocorticoids are naturally occurring hormones that prevent or suppress inflammation and immune responses when administered at pharmacological doses. At the molecular level, unbound glucocorticoids readily cross cell membranes and bind with high affinity to specific cytoplasmic receptors. Subsequent to binding, transcription and, ultimately, protein synthesis are affected. The result can include inhibition of leukocyte infiltration at the site of inflammation, interference in the function of mediators of inflammatory response, and suppression of humoral immune responses. The antiinflammatory actions of corticosteroids are thought to involve phospholipase A2 inhibitory proteins, collectively called lipocortins. Lipocortins, in turn, control the biosynthesis of potent mediators of inflammation such as prostaglandins and leukotrienes by inhibiting the release of the precursor molecule arachidonic acid. Further, corticosteroids inhibit eosinophil, basophil, and airway epithelial

WO 02/097032

PCT/US02/11152

cell function by regulation of cytokines that mediate the inflammatory response. They inhibit leukocyte infiltration at the site of inflammation, interfere in the function of mediators of the inflammatory response, and suppress the humoral immune response. Corticosteroids are used to treat allergies, asthma, arthritis, and skin conditions. Beclomethasone is a synthetic glucocorticoid that is used to treat steroid-dependent asthma, to relieve symptoms associated with allergic or nonallergic (vasomotor) rhinitis, or to prevent recurrent nasal polyps following surgical removal. The anti-inflammatory and vasoconstrictive effects of intranasal beclomethasone are 5000 times greater than those produced by hydrocortisone.

Expression profiling

Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

The discovery of new proteins associated with cell growth, differentiation, and death, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative disorders including cancer, developmental disorders, neurological disorders, autoimmune/inflammatory disorders, reproductive disorders, and disorders of the placenta, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of proteins associated with cell growth, differentiation, and death.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, proteins associated with cell growth, differentiation, and death, referred to collectively as "CGDD" and individually as "CGDD-1," "CGDD-2," "CGDD-3," "CGDD-4," "CGDD-5," "CGDD-6," "CGDD-7," "CGDD-8," "CGDD-9," "CGDD-10," "CGDD-11," "CGDD-12," "CGDD-13," "CGDD-14," "CGDD-15," "CGDD-16," "CGDD-17," "CGDD-18," "CGDD-19," "CGDD-20," and "CGDD-21." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid

WO 02/097032

PCT/US02/11152

sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the invention provides an isolated

5 polypeptide comprising the amino acid sequence of SEQ ID NO:1-21.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-21. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:22-42.

15 Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

25 The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a

WO 02/097032

PCT/US02/11152

polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90%

WO 02/097032

PCT/US02/11152

identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional CGDD, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional CGDD, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide

WO 02/097032

PCT/US02/11152

comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional CGDD, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

WO 02/097032

PCT/US02/11152

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting
 5 altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b)
 10 hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, iii) a polynucleotide
 15 having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, ii) a
 20 polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the
 25 amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30 BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank

WO 02/097032

PCT/US02/11152

homolog for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide sequences of the invention, along with allele frequencies in different human populations.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be

WO 02/097032

PCT/US02/11152

used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"CGDD" refers to the amino acid sequences of substantially purified CGDD obtained from
 5 any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of CGDD. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CGDD either by directly interacting with
 10 CGDD or by acting on components of the biological pathway in which CGDD participates.

An "allelic variant" is an alternative form of the gene encoding CGDD. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to
 15 allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding CGDD include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as CGDD or a
 20 polypeptide with at least one functional characteristic of CGDD. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding CGDD, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding CGDD. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of
 25 amino acid residues which produce a silent change and result in a functionally equivalent CGDD. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of CGDD is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and
 30 arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

WO 02/097032

PCT/US02/11152

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of CGDD. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CGDD either by directly interacting with CGDD or by acting on components of the biological pathway in which CGDD participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind CGDD polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The

WO 02/097032

PCT/US02/11152

nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH₂), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system.

5 Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) *J. Biotechnol.* 74:5-13.)

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) *Proc. Natl Acad. Sci. USA* 96:3606-3610).

10 The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

25 The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic CGDD, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

30 "Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a

WO 02/097032

PCT/US02/11152

given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding CGDD or fragments of CGDD may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

WO 02/097032

PCT/US02/11152

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

5 A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains
10 at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

15 "Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an
20 exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of CGDD or the polynucleotide encoding CGDD which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up
25 to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For
30 example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present

WO 02/097032

PCT/US02/11152

embodiments.

A fragment of SEQ ID NO:22-42 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:22-42, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:22-42 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:22-42 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:22-42 and the region of SEQ ID NO:22-42 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-21 is encoded by a fragment of SEQ ID NO:22-42. A fragment of SEQ ID NO:1-21 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-21. For example, a fragment of SEQ ID NO:1-21 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-21. The precise length of a fragment of SEQ ID NO:1-21 and the region of SEQ ID NO:1-21 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent

WO 02/097032

PCT/US02/11152

similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

15 Matrix: BLOSUM62
 Reward for match: 1
 Penalty for mismatch: -2
 Open Gap: 5 and Extension Gap: 2 penalties
 Gap x drop-off: 50
 Expect: 10
 Word Size: 11
 Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

30 Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

WO 02/097032

PCT/US02/11152

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62
Open Gap: 11 and Extension Gap: 1 penalties
Gap x drop-off: 50
Expect: 10
Word Size: 3
Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

WO 02/097032

PCT/US02/11152

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is

WO 02/097032

PCT/US02/11152

strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C₀t or R₀t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of CGDD which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of CGDD which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of CGDD. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of CGDD.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where

WO 02/097032

PCT/US02/11152

necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an CGDD may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of CGDD.

"Probe" refers to nucleic acid sequences encoding CGDD, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such

WO 02/097032

PCT/US02/11152

purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU
5 primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to
10 avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing
15 selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially
20 complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the
25 artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

30 Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated

WO 02/097032

PCT/US02/11152

regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing CGDD, nucleic acids encoding CGDD, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression

WO 02/097032

PCT/US02/11152

by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In one alternative, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), *supra*.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of

WO 02/097032

PCT/US02/11152

polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human proteins associated with cell growth, differentiation, and death (CGDD), the polynucleotides encoding CGDD, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative disorders including cancer, developmental disorders, neurological disorders, autoimmune/inflammatory disorders, reproductive disorders, and disorders of the placenta.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte

WO 02/097032

PCT/US02/11152

polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank homolog(s), along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are proteins associated with cell growth, differentiation, and death.

For example, SEQ ID NO:1 is 92% identical, from residue M1 to residue L1738, to murine ubiquitin-protein ligase E3-alpha (GenBank ID g3170887) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a putative zinc finger in N-recogin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from additional BLAST analyses provide further corroborative evidence that SEQ ID NO:1 is an ubiquitin protein ligase.

In an alternative example, For example, SEQ ID NO:3 is 88% identical, from residue M1 to residue D854, to murine ubiquitin-protein ligase Nedd4-2 (GenBank ID g12656270) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:3 also contains HECT (ubiquitin-transferase) and WW domains determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:3 is an ubiquitin-protein ligase.

WO 02/097032

PCT/US02/11152

In an alternative example, SEQ ID NO:5 is 100% identical, from residue M27 to residue D538, to human cisplatin resistance related protein CRR9p (GenBank ID g12248402) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 8.4e-281, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Data from additional BLAST analyses provide further corroborative evidence that SEQ ID NO:5 is an apoptosis-associated protein.

In an alternative example, SEQ ID NO:9 is 80% identical, from residue M1 to residue S710, to mouse RalGDS-like protein 3 (GenBank ID g8650435) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.5e-299, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:9 also contains a Ras association (RalGDS/AF-6) domain and a RasGEF domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from additional BLAST analyses using the PRODOM and DOMO databases provide further corroborative evidence that SEQ ID NO:9 is a guanine nucleotide dissociation factor.

In an alternative example, SEQ ID NO:12 is 77% identical, from residue A64 to residue Y365 and 100% identical from residue M1 to D109, to Sgt1 (GenBank ID g4809026) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.2e-121, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:12 also contains tetratricopeptide (TPR) domains from residue A45 to N78 and from residue S79 to T112, as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) TPR repeats are believed to mediate protein-protein interactions and are found in a number of proteins involved in mitosis. In addition, SPSCAN identifies a potential signal peptide from residue M1 through A68.

In an alternative example, SEQ ID NO:13 is 100% identical, from residue M1 to residue K365, to human proto-oncogene Wnt-5A (GenBank ID g348918) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.5e-205, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:13 also contains a wnt-1 family domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:13 is a wnt-1 family protein, a member of the wnt

WO 02/097032

PCT/US02/11152

family of secreted glycoproteins.

In an alternative example, SEQ ID NO:16 is 50% identical, from residue A18 to residue F1014, to human cyclin-E binding protein 1 (GenBank ID g6630609) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.6e-252, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:16 also contains a HECT (ubiquitin transferase) domain, and a regulator of chromosome condensation (RCC1) protein domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:16 is a cyclin-binding protein.

In an alternative example, SEQ ID NO:17 is 99% identical, from residue M1 to residue S1462, to a human cyclophilin-related protein (GenBank ID g5923891) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:17 also contains a cyclophilin-type peptidyl-prolyl cis-trans isomerase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:17 is a cyclophilin-related protein.

In an alternative example, SEQ ID NO:19 is 34% identical, from residue K3 to residue S175, and 26% identical, from residue R40 to Q327, to human apoptotic protease activating factor 1 (GenBank ID g2330015) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 8.3e-21, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:19 also contains a SAM domain and G-protein beta WD-40 repeats as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:19 contains multiple beta G-protein WD-40 signatures similarly to Apaf-1.

SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6-8, SEQ ID NO:10-11, SEQ ID NO:14-15, SEQ ID NO:18, and SEQ ID NO:20-21 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-21 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were

WO 02/097032

PCT/US02/11152

assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:22-42 or that distinguish between SEQ ID NO:22-42 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_{XXXXXX}_{N₁}_{N₂}_{YYYY}_{N₃}_{N₄} represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYY is the number of the prediction generated by the algorithm, and N_{1,2,3,4}, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FLXXXXX_gAAAAA_gBBBBB_1_N is a "stretched" sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM,"

WO 02/097032

PCT/US02/11152

"NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, gBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses CGDD variants. A preferred CGDD variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the CGDD amino acid sequence, and which contains at least one functional or structural characteristic of CGDD.

The invention also encompasses polynucleotides which encode CGDD. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:22-42, which encodes CGDD. The polynucleotide sequences of SEQ ID NO:22-42, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

WO 02/097032

PCT/US02/11152

The invention also encompasses a variant of a polynucleotide sequence encoding CGDD. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding CGDD. A particular aspect of the invention encompasses a variant of a
5 polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:22-42 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:22-42. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CGDD.

10 In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding CGDD. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding CGDD, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%,
15 or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding CGDD over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding CGDD. For example, a polynucleotide comprising a sequence of SEQ ID NO:42
20 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:41. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CGDD.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding CGDD, some bearing minimal
25 similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring CGDD, and all such variations are to be considered as
30 being specifically disclosed.

Although nucleotide sequences which encode CGDD and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring CGDD under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding CGDD or

WO 02/097032

PCT/US02/11152

its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding CGDD and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode CGDD and CGDD derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding CGDD or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:22-42 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) *Short Protocols in Molecular Biology*, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) *Molecular Biology and Biotechnology*, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding CGDD may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences,

WO 02/097032

PCT/US02/11152

such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:S186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060.) Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

WO 02/097032

PCT/US02/11152

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode CGDD may be cloned in recombinant DNA molecules that direct expression of CGDD, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express CGDD.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter CGDD-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULAR BREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of CGDD, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding CGDD may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, CGDD itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp.

WO 02/097032

PCT/US02/11152

55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of CGDD, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, *supra*, pp. 28-53.)

10 In order to express a biologically active CGDD, the nucleotide sequences encoding CGDD or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding CGDD. Such elements may vary in their strength and specificity. Specific initiation signals 15 may also be used to achieve more efficient translation of sequences encoding CGDD. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding CGDD and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may 20 be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. 25 Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding CGDD and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning. A Laboratory 30 Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding CGDD. These include, but are not limited to, microorganisms such as bacteria transformed

WO 02/097032

PCT/US02/11152

with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or
 5 animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and
 10 Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994)
 15 Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding CGDD. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding CGDD can be achieved using a
 20 multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding CGDD into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for
 in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of
 25 nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of CGDD are needed, e.g. for the production of antibodies, vectors which direct high level expression of CGDD may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of CGDD. A number of vectors
 30 containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra;

WO 02/097032

PCT/US02/11152

Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of CGDD. Transcription of sequences encoding CGDD may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding CGDD may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses CGDD in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of CGDD in cell lines is preferred. For example, sequences encoding CGDD can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

WO 02/097032

PCT/US02/11152

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apv* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β glucuronidase and its substrate β -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding CGDD is inserted within a marker gene sequence, transformed cells containing sequences encoding CGDD can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding CGDD under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding CGDD and that express CGDD may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of CGDD using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CGDD is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See,

WO 02/097032

PCT/US02/11152

e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

5 A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding CGDD include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding CGDD, or any fragments thereof, may be cloned into a vector
10 for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for
15 ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding CGDD may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence
20 and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CGDD may be designed to contain signal sequences which direct secretion of CGDD through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of
25 the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture
30 Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding CGDD may be ligated to a heterologous sequence resulting in translation of a

WO 02/097032

PCT/US02/11152

fusion protein in any of the aforementioned host systems. For example, a chimeric CGDD protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of CGDD activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity
 5 matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion
 10 proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the CGDD encoding sequence and the heterologous protein sequence, so that CGDD may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially
 15 available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled CGDD may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid
 20 precursor, for example, ³⁵S-methionine.

CGDD of the present invention or fragments thereof may be used to screen for compounds that specifically bind to CGDD. At least one and up to a plurality of test compounds may be screened for specific binding to CGDD. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

25 In one embodiment, the compound thus identified is closely related to the natural ligand of CGDD, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which CGDD binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the
 30 compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express CGDD, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing CGDD or cell membrane fractions which contain CGDD are then contacted

WO 02/097032

PCT/US02/11152

with a test compound and binding, stimulation, or inhibition of activity of either CGDD or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with CGDD, either in solution or affixed to a solid support, and detecting the binding of CGDD to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

CGDD of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of CGDD. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for CGDD activity, wherein CGDD is combined with at least one test compound, and the activity of CGDD in the presence of a test compound is compared with the activity of CGDD in the absence of the test compound. A change in the activity of CGDD in the presence of the test compound is indicative of a compound that modulates the activity of CGDD. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising CGDD under conditions suitable for CGDD activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of CGDD may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding CGDD or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) *Science* 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) *Clin. Invest.* 97:1999-2002; Wagner, K.U. et al. (1997) *Nucleic Acids Res.* 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and

WO 02/097032

PCT/US02/11152

the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding CGDD may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding CGDD can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding CGDD is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress CGDD, e.g., by secreting CGDD in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of CGDD and proteins associated with cell growth, differentiation, and death. In addition, examples of tissues expressing CGDD are breast cancer, PBMC cells, and brain cingulate tissue, and also can be found in Table 6. Therefore, CGDD appears to play a role in cell proliferative disorders including cancer, developmental disorders, neurological disorders, autoimmune/inflammatory disorders, reproductive disorders, and disorders of the placenta. In the treatment of disorders associated with increased CGDD expression or activity, it is desirable to decrease the expression or activity of CGDD. In the treatment of disorders associated with decreased CGDD expression or activity, it is desirable to increase the expression or activity of CGDD.

Therefore, in one embodiment, CGDD or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CGDD. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract,

WO 02/097032

PCT/US02/11152

heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Strausler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis,

WO 02/097032

PCT/US02/11152

hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, 5 fungal, parasitic, protozoal, and helminthic infections, and trauma; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, ectopic pregnancy, teratogenesis; cancer of the breast, fibrocystic breast disease, galactorrhea; a 10 disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, gynecomastia, hypergonadotropic and hypogonadotropic hypogonadism, pseudohermaphroditism, azoospermia, premature ovarian failure, acrosin deficiency, delayed puberty, retrograde ejaculation 15 and anejaculation, haemangioblastomas, cystsphaechromocytomas, paraganglioma, cystadenomas of the epididymis, and endolymphatic sac tumors; and a disorder of the placenta such as preeclampsia, choriocarcinoma, abruptio placentae, placenta previa, placental or maternal floor infarction, placenta accreta, increta, and percreta, extrachorial placentas, chorangioma, chorangiosis, chronic villitis, placental villous endema, widespread fibrosis of the terminal villi, intervillous thrombi, hemorrhagic 20 endovasculitis, erythroblastosis fetalis, and nonimmune fetal hydrops.

In another embodiment, a vector capable of expressing CGDD or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CGDD including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified CGDD in 25 conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CGDD including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of CGDD may be administered to a subject to treat or prevent a disorder associated with decreased expression or 30 activity of CGDD including, but not limited to, those listed above.

In a further embodiment, an antagonist of CGDD may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CGDD. Examples of such disorders include, but are not limited to, those cell proliferative disorders including cancer,

WO 02/097032

PCT/US02/11152

developmental disorders, neurological disorders, autoimmune/inflammatory disorders, reproductive disorders, and disorders of the placenta described above. In one aspect, an antibody which specifically binds CGDD may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express CGDD.

5 In an additional embodiment, a vector expressing the complement of the polynucleotide encoding CGDD may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CGDD including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate
10 therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

15 An antagonist of CGDD may be produced using methods which are generally known in the art. In particular, purified CGDD may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind CGDD. Antibodies to CGDD may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and
20 fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

25 For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with CGDD or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such
30 as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to

WO 02/097032

PCT/US02/11152

CGDD have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of CGDD amino acids may be fused with those of another protein, such as KLH, and antibodies to the

5 chimeric molecule may be produced.

Monoclonal antibodies to CGDD may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and

10 Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc.*

15 *Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce CGDD-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton,

20 D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299.)

25 Antibody fragments which contain specific binding sites for CGDD may also be generated. For example, such fragments include, but are not limited to, F(ab')₂ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D.

30 et al. (1989) *Science* 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such

WO 02/097032

PCT/US02/11152

immunoassays typically involve the measurement of complex formation between CGDD and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CGDD epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

5 Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for CGDD. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of CGDD-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for
10 multiple CGDD epitopes, represents the average affinity, or avidity, of the antibodies for CGDD. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular CGDD epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the CGDD-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a
15 ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of CGDD, preferably in active form, from the antibody (Catty, D. (1988) *Antibodies, Volume I: A Practical Approach*, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) *A Practical Guide to Monoclonal Antibodies*, John Wiley & Sons, New York NY).

20 The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of CGDD-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for
25 antibody quality and usage in various applications, are generally available. (See, e.g., Catty, *supra*, and Coligan et al. *supra*.)

In another embodiment of the invention, the polynucleotides encoding CGDD, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA,
30 RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding CGDD. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding CGDD. (See, e.g., Agrawal, S., ed. (1996) *Antisense Therapeutics*, Humana Press Inc., Totawa NJ.)

WO 02/097032

PCT/US02/11152

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding CGDD may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in CGDD expression or regulation causes disease, the expression of CGDD from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in CGDD are treated by constructing mammalian expression vectors encoding CGDD and introducing

WO 02/097032

PCT/US02/11152

these vectors by mechanical means into CGDD-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.*

5 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J.-L. and H. R  cipon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of CGDD include, but are not limited to, the pCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA),
 10 and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). CGDD may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-5551; Gossen, M. et al. (1995) *Science* 268:1766-1769; Rossi, F.M.V. and H.M. Blau
 15 (1998) *Curr. Opin. Biotechnol.* 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and F1ND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding CGDD from a normal individual.

20 Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al.
 25 (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to CGDD expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding CGDD under the control of an independent promoter or the retrovirus long
 30 terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc.*

WO 02/097032

PCT/US02/11152

Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference.

Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding CGDD to cells which have one or more genetic abnormalities with respect to the expression of CGDD. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding CGDD to target cells which have one or more genetic abnormalities with respect to the expression of CGDD. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing CGDD to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 69:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92

WO 02/097032

PCT/US02/11152

which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

10 In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding CGDD to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, 15 resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for CGDD into the alphavirus genome in place of the capsid-coding region results in the production of a large number of CGDD-coding RNAs and the synthesis of high levels of CGDD in vector transduced cells. While 20 alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of CGDD into a variety of cell types. The specific transduction of a subset of 25 cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can 30 be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have

WO 02/097032

PCT/US02/11152

been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

5 Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding CGDD.

10 Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. 20 Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding CGDD. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

25 RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, 30 and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding CGDD. Compounds

WO 02/097032

PCT/US02/11152

which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased CGDD expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding CGDD may be therapeutically useful, and in the treatment of disorders associated with decreased CGDD expression or activity, a compound which specifically promotes expression of the polynucleotide encoding CGDD may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding CGDD is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding CGDD are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding CGDD. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al.

WO 02/097032

PCT/US02/11152

(1997) U.S. Patent No. 5,686,242; Brucis, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient.

- 5 Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

- Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and
10 monkeys.

- An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's
15 Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of CGDD, antibodies to CGDD, and mimetics, agonists, antagonists, or inhibitors of CGDD.

- The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical,
20 sublingual, or rectal means.

- Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and
25 proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

- Compositions suitable for use in the invention include compositions wherein the active
30 ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising CGDD or fragments thereof. For example, liposome preparations

WO 02/097032

PCT/US02/11152

containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, CGDD or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example CGDD or fragments thereof, antibodies of CGDD, and agonists, antagonists or inhibitors of CGDD, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their

WO 02/097032

PCT/US02/11152

inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind CGDD may be used for the
 5 diagnosis of disorders characterized by expression of CGDD, or in assays to monitor patients being
 treated with CGDD or agonists, antagonists, or inhibitors of CGDD. Antibodies useful for diagnostic
 purposes may be prepared in the same manner as described above for therapeutics. Diagnostic
 assays for CGDD include methods which utilize the antibody and a label to detect CGDD in human
 body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification,
 10 and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of
 reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring CGDD, including ELISAs, RIAs, and FACS, are known
 in the art and provide a basis for diagnosing altered or abnormal levels of CGDD expression. Normal
 or standard values for CGDD expression are established by combining body fluids or cell extracts
 15 taken from normal mammalian subjects, for example, human subjects, with antibodies to CGDD under
 conditions suitable for complex formation. The amount of standard complex formation may be
 quantitated by various methods, such as photometric means. Quantities of CGDD expressed in
 subject, control, and disease samples from biopsied tissues are compared with the standard values.
 Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding CGDD may be used for
 diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences,
 complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect
 and quantify gene expression in biopsied tissues in which expression of CGDD may be correlated with
 disease. The diagnostic assay may be used to determine absence, presence, and excess expression of
 20 CGDD, and to monitor regulation of CGDD levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide
 sequences, including genomic sequences, encoding CGDD or closely related molecules may be used
 to identify nucleic acid sequences which encode CGDD. The specificity of the probe, whether it is
 made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a
 30 conserved motif, and the stringency of the hybridization or amplification will determine whether the
 probe identifies only naturally occurring sequences encoding CGDD, allelic variants, or related
 sequences.

Probes may also be used for the detection of related sequences, and may have at least 50%

WO 02/097032

PCT/US02/11152

sequence identity to any of the CGDD encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:22-42 or from genomic sequences including promoters, enhancers, and introns of the CGDD gene.

Means for producing specific hybridization probes for DNAs encoding CGDD include the
 5 cloning of polynucleotide sequences encoding CGDD or CGDD derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ³²P or ³⁵S, or by enzymatic labels,
 10 such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding CGDD may be used for the diagnosis of disorders associated with expression of CGDD. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal
 15 hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder such
 20 as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure
 25 disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular
 30 atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome,

WO 02/097032

PCT/US02/11152

fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, ectopic pregnancy, teratogenesis; cancer of the breast, fibrocystic breast disease, galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, gynecomastia, hypergonadotropic and hypogonadotropic hypogonadism, pseudohermaphroditism, azoospermia, premature ovarian failure, acrosin deficiency, delayed puberty, retrograde ejaculation and anejaculation, haemangioblastomas, cystsphaechromocytomas, paraganglioma, cystadenomas of the epididymis, and endolymphatic sac tumors; and a disorder of the

WO 02/097032

PCT/US02/11152

placenta such as preeclampsia, choriocarcinoma, abruptio placentae, placenta previa, placental or maternal floor infarction, placenta accreta, increta, and percreta, extrachorial placentas, chorangioma, chorangiosis, chronic villitis, placental villous endema, widespread fibrosis of the terminal villi, intervillous thrombi, hemorrhagic endovasculitis, erythroblastosis fetalis, and nonimmune fetal

5 hydrops. The polynucleotide sequences encoding CGDD may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multifomat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered CGDD expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding CGDD may be useful in assays that

10 detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding CGDD may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control

15 sample then the presence of altered levels of nucleotide sequences encoding CGDD in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of

20 CGDD, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding CGDD, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified

25 polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the

30 patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or

WO 02/097032

PCT/US02/11152

overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

5 Additional diagnostic uses for oligonucleotides designed from the sequences encoding CGDD may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding CGDD, or a fragment of a polynucleotide complementary to the polynucleotide encoding
10 CGDD, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding CGDD may be used to detect single nucleotide polymorphisms (SNPs). SNPs are
15 substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding CGDD are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal
20 tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP
25 (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the
30 high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis,

WO 02/097032

PCT/US02/11152

sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in

5 N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations

10 and their migrations. (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641.)

Methods which may also be used to quantify the expression of CGDD include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C.

15 et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the

20 polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor

25 progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic

30 profile.

In another embodiment, CGDD, fragments of CGDD, or antibodies specific for CGDD may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

WO 02/097032

PCT/US02/11152

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at
5 a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present
10 invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

15 Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity
20 (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression
25 provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not
30 necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed

WO 02/097032

PCT/US02/11152

gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for CGDD to quantify the levels of CGDD expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-

WO 02/097032

PCT/US02/11152

111; Mendoz, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

5 Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Scilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which
10 alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of
15 each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

20 In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the
25 two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) *Proc. Natl. Acad. Sci. USA*
30 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding CGDD may be used

WO 02/097032

PCT/US02/11152

to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding CGDD on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, CGDD, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a

WO 02/097032

PCT/US02/11152

solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between CGDD and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with CGDD, or fragments thereof, and washed. Bound CGDD is then detected by methods well known in the art. Purified CGDD can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding CGDD specifically compete with a test compound for binding CGDD. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with CGDD.

In additional embodiments, the nucleotide sequences which encode CGDD may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/286,820, U.S. Ser. No. 60/293,727, U.S. Ser. No. 60/283,294, U.S. Ser. No. 60/282,110, U.S. Ser. No. 60/287,228, U.S. Ser. No. 60/291,846, U.S. Ser. No. 60/291,662, U.S. Ser. No. 60/295,340, U.S. Ser. No. 60/295,263, and U.S. Ser. No. 60/349,705, are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incye cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incye Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of

WO 02/097032

PCT/US02/11152

denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

5 Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)⁺ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the

10 POLY(A)⁺PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPTM plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*,
 15 units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs
 20 were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were
 25 transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5a, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using
 30 at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1

WO 02/097032

PCT/US02/11152

ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows.

Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic

WO 02/097032

PCT/US02/11152

Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) *Curr. Opin. Struct. Biol.* 6:361-365.)

The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER.

The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences.

- 5 Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Couped, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive
- 10 the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM,
- 15 INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program
- 20 (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of

- 25 which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and

- 30 polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:22-42. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

WO 02/097032

PCT/US02/11152

Putative proteins associated with cell growth, differentiation, and death were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpr1 and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode proteins associated with cell growth, differentiation, and death, the encoded polypeptides were analyzed by querying against PFAM models for proteins associated with cell growth, differentiation, and death. Potential proteins associated with cell growth, differentiation, and death were also identified by homology to Incyte cDNA sequences that had been annotated as proteins associated with cell growth, differentiation, and death. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpr1 public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on

WO 02/097032

PCT/US02/11152

more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the *genpept* and *gbpri* public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from *genpept*. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of CGDD Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:22-42 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:22-42 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences

WO 02/097032

PCT/US02/11152

had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:26 was mapped to chromosome 3 within the interval from 63.30 to 77.40 centiMorgans.

15 VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

20 Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$25 \quad \frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length}(\text{Seq. 1}), \text{length}(\text{Seq. 2}) \}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair

WO 02/097032

PCT/US02/11152

(HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding CGDD are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding CGDD. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of CGDD Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68 °C to about 72 °C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one

WO 02/097032

PCT/US02/11152

extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+} , $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μ l to 10 μ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviHI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was

WO 02/097032

PCT/US02/11152

quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM

5 BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in CGDD Encoding

10 Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:22-42 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of 15 basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify 20 errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

25 Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezuelan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The 30 Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed

WO 02/097032

PCT/US02/11152

no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:22-42 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base
 5 pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ - ^{32} P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a
 10 SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon
 15 membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

20 XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schemm (1999), *supra*).
 25 Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schemm, M. et al. (1995) Science
 30 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be

WO 02/097032

PCT/US02/11152

selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection.

After hybridization, nonhybridized nucleotides from the biological sample are removed, and a
 5 fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

10 Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M
 15 dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and
 20 incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended
 25 in 14 μ l 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are
 30 amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water

WO 02/097032

PCT/US02/11152

washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

5 Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average concentration of 100 ng/ μ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene).

10 Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

15 Hybridization reactions contain 9 μ l of sample mixture consisting of 0.2 μ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 20 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an 25 Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a 30 resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate

WO 02/097032

PCT/US02/11152

filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

5 The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, 10 are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC 15 computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

20 A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

Expression

25 Normal breast cell lines are obtained as follows. Primary mammary gland cells are isolated from a donor with fibrocystic breast disease. Alternatively, primary breast epithelial cells are isolated from a normal donor. Breast carcinoma cells are derived in vitro from cells emigrating from a tumor. Normal and various stages of tumorigenic breast cell lines were purchased from American Type Culture Collection (ATCC), (Manassas, VA).

30 For example, SEQ ID NO:34 showed differential expression in cancer cell lines or tumorous tissue versus non-cancerous cell lines or tissues as determined by microarray analysis. The expression of CGDD-13 was increased by at least three fold in a breast tumor cell line that was harvested from a donor with an early stage of tumor progression.

WO 02/097032

PCT/US02/11152

In an alternative example, SEQ ID NO:37 showed differential expression in inflammatory responses as determined by microarray analysis. The expression of SEQ ID NO:37 was increased by at least two fold in PBMCs treated with LPS relative to untreated PBMCs. Therefore, SEQ ID NO:37 is useful in diagnostic assays for inflammatory responses.

In addition, SEQ ID NO:37 showed differential expression in non-malignant mammary epithelial cells versus various breast carcinoma lines as determined by microarray analysis. The expression of SEQ ID NO:37 was decreased by at least two fold in the breast carcinoma lines relative to non-malignant mammary epithelial cells. Therefore, SEQ ID NO:37 is useful in diagnostic assays for detection of breast cancer.

In an alternative example, SEQ ID NO:41 showed differential expression in brain cingulate from a patient with Alzheimer's disease compared to matched microscopically normal tissue from the same donor as determined by microarray analysis. The expression of CGDD-19 was increased in cingulate tissue with Alzheimer's disease. Therefore, SEQ ID NO:41 is useful in diagnostic assays for neurological disorders, particularly Alzheimer's disease.

XII. Complementary Polynucleotides

Sequences complementary to the CGDD-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring CGDD. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of CGDD. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the CGDD-encoding transcript.

XIII. Expression of CGDD

Expression and purification of CGDD is achieved using bacterial or virus-based expression systems. For expression of CGDD in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express CGDD upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of CGDD in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus

WO 02/097032

PCT/US02/11152

(AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding CGDD by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (SF9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, CGDD is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from CGDD at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10 and 16). Purified CGDD obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, and XIX, where applicable.

XIV. Functional Assays

CGDD function is assessed by expressing the sequences encoding CGDD at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate

WO 02/097032

PCT/US02/11152

the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of CGDD on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding CGDD and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding CGDD and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of CGDD Specific Antibodies

CGDD substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the CGDD amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-CGDD activity by, for example, binding the peptide or CGDD to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat

WO 02/097032

PCT/US02/11152

anti-rabbit IgG.

XVI. Purification of Naturally Occurring CGDD Using Specific Antibodies

Naturally occurring or recombinant CGDD is substantially purified by immunoaffinity chromatography using antibodies specific for CGDD. An immunoaffinity column is constructed by covalently coupling anti-CGDD antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing CGDD are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of CGDD (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/CGDD binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and CGDD is collected.

XVII. Identification of Molecules Which Interact with CGDD

CGDD, or biologically active fragments thereof, are labeled with ¹²⁵I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled CGDD, washed, and any wells with labeled CGDD complex are assayed. Data obtained using different concentrations of CGDD are used to calculate values for the number, affinity, and association of CGDD with the candidate molecules.

Alternatively, molecules interacting with CGDD are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

CGDD may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of CGDD Activity

CGDD activity is demonstrated by measuring the induction of terminal differentiation or cell cycle progression when CGDD is expressed at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies, Gaithersburg, MD) and PCR 3.1 (Invitrogen, Carlsbad, CA), both of which contain the cytomegalovirus promoter. 5-10 µg of recombinant vector are transiently transfected into

WO 02/097032

PCT/US02/11152

a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 μ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant

5 vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP) (Clontech, Palo Alto, CA), CD64, or a CD64-GFP fusion protein. Flow cytometry detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell cycle progression or terminal differentiation. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward

10 light scatter and 90 degree side light scatter; up or down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow Cytometry,

15 Oxford, New York, NY.

Alternatively, an in vitro assay for CGDD activity measures the transformation of normal human fibroblast cells overexpressing antisense CGDD RNA (Garkavtsev, I. and Riabowol, K. (1997) Mol. Cell Biol. 17:2014-2019). cDNA encoding CGDD is subcloned into the pLNCX retroviral vector to enable expression of antisense CGDD RNA. The resulting construct is transfected into the

20 ecotropic BOSC23 virus-packaging cell line. Virus contained in the BOSC23 culture supernatant is used to infect the amphotropic CAK8 virus-packaging cell line. Virus contained in the CAK8 culture supernatant is used to infect normal human fibroblast (Hs68) cells. Infected cells are assessed for the following quantifiable properties characteristic of transformed cells: growth in culture to high density associated with loss of contact inhibition, growth in suspension or in soft agar, formation of colonies or

25 foci, lowered serum requirements, and ability to induce tumors when injected into immunodeficient mice. The activity of CGDD is proportional to the extent of transformation of Hs68 cells.

Alternatively, CGDD can be expressed in a mammalian cell line by transforming the cells with a eukaryotic expression vector encoding CGDD. Eukaryotic expression vectors are commercially available, and the techniques to introduce them into cells are well known to those skilled in the art. To

30 assay the cellular localization of CGDD, cells are fractionated as described by Jiang H. P. et al. (1992; Proc. Natl. Acad. Sci. 89: 7856-7860). Briefly, cells pelleted by low-speed centrifugation are resuspended in buffer (10 mM TRIS-HCl, pH 7.4/ 10 mM NaCl/ 3 mM MgCl₂/ 5 mM EDTA with 10 ug/ml aprotinin, 10 ug/ml leupeptin, 10 ug/ml pepstatin A, 0.2 mM phenylmethylsulfonyl fluoride) and

WO 02/097032

PCT/US02/11152

homogenized. The homogenate is centrifuged at 600 x g for 5 minutes. The particulate and cytosol fractions are separated by ultracentrifugation of the supernatant at 100,000 x g for 60 minutes. The nuclear fraction is obtained by resuspending the 600 x g pellet in sucrose solution (0.25 M sucrose/ 10 mM TRIS-HCl, pH 7.4/ 2 mM MgCl₂) and recentrifuged at 600 x g. Equal amounts of protein from each fraction are applied to an SDS/10% polyacrylamide gel and blotted onto membranes. Western blot analysis is performed using CGDD anti-serum. The localization of CGDD is assessed by the intensity of the corresponding band in the nuclear fraction relative to the intensity in the other fractions. Alternatively, the presence of CGDD in cellular fractions is examined by fluorescence microscopy using a fluorescent antibody specific for CGDD.

Alternatively, CGDD activity may be demonstrated as the ability to interact with its associated Ras superfamily protein, in an *in vitro* binding assay. The candidate Ras superfamily proteins are expressed as fusion proteins with glutathione S-transferase (GST), and purified by affinity chromatography on glutathione-Sepharose. The Ras superfamily proteins are loaded with GDP by incubating 20 mM Tris buffer, pH 8.0, containing 100 mM NaCl, 2 mM EDTA, 5 mM MgCl₂, 0.2 mM DTT, 100 μ M AMP-PNP and 10 μ M GDP at 30°C for 20 minutes. CGDD is expressed as a FLAG fusion protein in a baculovirus system. Extracts of these baculovirus cells containing CGDD-FLAG fusion proteins are precleared with GST beads, then incubated with GST-Ras superfamily fusion proteins. The complexes formed are precipitated by glutathione-Sepharose and separated by SDS-polyacrylamide gel electrophoresis. The separated proteins are blotted onto nitrocellulose membranes and probed with commercially available anti-FLAG antibodies. CGDD activity is proportional to the amount of CGDD-FLAG fusion protein detected in the complex.

Alternatively, as demonstrated by Li and Cohen (Li, L. and S.N. Cohen (1995) Cell 85:319-329), the ability of CGDD to suppress tumorigenesis can be measured by designing an antisense sequence to the 5' end of the gene and transfecting NIH 3T3 cells with a vector transcribing this sequence. The suppression of the endogenous gene will allow transformed fibroblasts to produce clumps of cells capable of forming metastatic tumors when introduced into nude mice.

Alternatively, an assay for CGDD activity measures the effect of injected CGDD on the degradation of maternal transcripts. Procedures for oocyte collection from Swiss albino mice, injection, and culture are as described in Stutz (supra). A decrease in the degradation of maternal RNAs as compared to control oocytes is indicative of CGDD activity. In the alternative, CGDD activity is measured as the ability of purified CGDD to bind to RNase as measured by the assays described in Example XVII.

Alternatively, an assay for CGDD activity measures syncytium formation in COS cells

WO 02/097032

PCT/US02/11152

transfected with an CGDD expression plasmid, using the two-component fusion assay described in Mi (supra). This assay takes advantage of the fact that human interleukin 12 (IL-12) is a heterodimer comprising subunits with molecular weights of 35 kD (p35) and 40 kD (p40). COS cells transfected with expression plasmids carrying the gene for p35 are mixed with COS cells cotransfected with
 5 expression plasmids carrying the genes for p40 and CGDD. The level of IL-12 activity in the resulting conditioned medium corresponds to the activity of CGDD in this assay. Syncytium formation may also be measured by light microscopy (Mi et al. supra).

An alternative assay for CGDD activity measures cell proliferation as the amount of newly initiated DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding
 10 CGDD is transfected into quiescent 3T3 cultured cells using methods well known in the art. The transiently transfected cells are then incubated in the presence of [³H]thymidine or a radioactive DNA precursor such as [α -³²P]ATP. Where applicable, varying amounts of CGDD ligand are added to the transfected cells. Incorporation of [³H]thymidine into acid-precipitable DNA is measured over an appropriate time interval, and the amount incorporated is directly proportional to the amount of newly
 15 synthesized DNA and CGDD activity.

Alternatively, CGDD activity is measured by the cyclin-ubiquitin ligation assay (Townsend, F.M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2362-2367). The reaction contains in a volume of 10 μ l, 40 mM Tris.HCl (pH 7.6), 5 mM Mg Cl₂, 0.5 mM ATP, 10 mM phosphocreatine, 50 μ g of creatine phosphokinase/ml, 1 mg reduced carboxymethylated bovine serum albumin/ml, 50 μ M
 20 ubiquitin, 1 μ M ubiquitin aldehyde, 1-2 pmol ¹²⁵I-labeled cyclin B, 1 pmol E1, 1 μ M okadaic acid, 10 μ g of protein of M-phase fraction 1A (containing active E3-C and essentially free of E2-C), and varying amounts of CGDD. The reaction is incubated at 18 °C for 60 minutes. Samples are then separated by electrophoresis on an SDS polyacrylamide gel. The amount of ¹²⁵I- cyclin-ubiquitin formed is quantified by PHOSPHORIMAGER analysis. The amount of cyclin-ubiquitin formation is
 25 proportional to the activity of CGDD in the reaction.

Alternatively, an assay for CGDD activity uses radiolabeled nucleotides, such as [α -³²P]ATP, to measure either the incorporation of radiolabel into DNA during DNA synthesis, or fragmentation of DNA that accompanies apoptosis. Mammalian cells are transfected with plasmid containing cDNA encoding CGDD by methods well known in the art. Cells are then incubated with radiolabeled
 30 nucleotide for various lengths of time. Chromosomal DNA is collected, and radioactivity is detected using a scintillation counter. Incorporation of radiolabel into chromosomal DNA is proportional to the degree of stimulation of the cell cycle. To determine if CGDD promotes apoptosis, chromosomal DNA is collected as above, and analyzed using polyacrylamide gel electrophoresis, by methods well

WO 02/097032

PCT/US02/11152

known in the art. Fragmentation of DNA is quantified by comparison to untransfected control cells, and is proportional to the apoptotic activity of CGDD.

Alternatively, cyclophilin activity of CGDD is measured using a chymotrypsin-coupled assay to measure the rate of cis to trans interconversion (Fischer, G., Bang, H., and Mech, C. (1984) Biomed. Biochim. Acta 43: 1101-1111). The chymotrypsin is used to estimate the trans-substrate cleavage activity at Xaa-Pro peptide bonds, wherein the rate constant for the cis to trans isomerization can be obtained by measuring the rate constant of the substrate hydrolysis at the slow phase. Samples are incubated in the presence or absence of the immunosuppressant drugs CsA or FK506, reactions initiated by addition of chymotrypsin, and the fluorescent reaction measured. The enzymatic rate constant is calculated from the equation $k_{app} = k_{R20} + k_{cat}$, wherein first order kinetics are displayed, and where one unit of PPIase activity is defined as k_{cat} (s^{-1}).

A fluorescence monitoring assay for detecting activated Ras using RRP22 is as follows. The RRP22 binding domain (RRP22BD) of c-Raf1 (a kinase activated during reentry into meiosis) is synthesized from two unprotected peptide segments by native chemical ligation. Two fluorescent amino acids with structures based on the nitrobenz-2-oxa-1,3-diazole and coumaryl chromophores are incorporated close to the RRP22BD/RRP22-GTP binding surface followed by introduction of a C-terminal tag consisting of His(6). The K_D values for binding of the site-specifically modified proteins to Ras-GTP are compared to that of wild-type RBD. Ras-GTP is detected within the 100 nM range by immobilization of C-terminal His(6) tag-modified fluorescent RBD onto Ni-NTA-coated surfaces. Ras-GDP does not bind to the immobilized RBD, thus allowing discrimination between inactive and activated Ras (Becker, C. F. (2001) Chem. Biol. 8:243-252).

CGDD is assayed for Ras binding by the method of Vavvas *et al.* (*supra*). CGDD is expressed as a GST fusion protein and the GST-CGDD fusion is incubated with Ras in the presence of either GTPγS or GDPβS. Glutathione-SEPHAROSE beads (APB) are added to recover the GST-CGDD fusion and GST-CGDD-Ras complexes from solution. Proteins are eluted from the glutathione-SEPHAROSE beads with SDS sample buffer and separated by SDS-PAGE. Following electrophoresis, proteins are transferred to a PVDF membrane (APB) and probed for Ras with monoclonal anti-Ras antibodies.

Regulation of Wnt5a by cell-to-cell contacts is shown by adding various metabolic agents that selectively block protein tyrosine kinases (genistein) or cytochalasin D to HB2, a normal breast epithelial cell line. Cytoskeleton reorganization following cytochalasin D treatment causes an induction of Wnt5a, which is associated with changes in cell morphology. Cancer cell lines treated with cytochalasin D show no changes in cell morphology nor Wnt5a induction (Jonsson, M. *et al.* (1998)

WO 02/097032

PCT/US02/11152

Br. J. Cancer 78:430-438).

XIX. CGDD Binding Assays

A quantitative immunoassay for the CGDD cyclophilin measures its affinity for stereospecific binding to the immunosuppressant drug cyclosporin (Quesniaux, V.F., et al. (1987) Eur. J. Immunol. 17: 1359-1365). In this assay, the cyclophilin-cyclosporin complex is coated on a solid phase, with binding detected using anti-cyclophilin rabbit antiserum enhanced by an antiglobulin-enzyme conjugate. Complexing of the CGDD immunophilin, cyclophilin, with the immunosuppressant drug cyclosporin at critical residues facilitates immunosuppressant activity, such as that which occurs during tumorigenesis.

A binding assay developed to measure the non-covalent binding between FKBP's and immunosuppressant drugs in the gas phase utilizes electrospray ionization mass spectrometry (Trepanier, D.J., et al. (1999) Ther. Drug Monit. 21: 274-280). In electrospray ionization, ions are generated by creating a fine spray of highly charged droplets in the presence of a strong electric field; as the droplet decreases in size, the charge density on the surface increases. Ions are electrostatically directed into a mass analyzer, where ions of opposite charge are generated in spatially separate sources and then swept into capillary inlets where the flows are merged and where reactions occur. By comparing the charge states of bound versus unbound FKBP/immunosuppressive drug complexes, relative binding affinities can be established and correlated with *in vitro* binding and immunosuppressive activity.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

WO 02/097032

PCT/US02/11152

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	CA2 Reagents
1351608	1	1351608CD1	22	1351608CB1	
4259314	2	4259314CD1	23	4259314CB1	2747431CA2, 3421573CA2
3660046	3	3660046CD1	24	3660046CB1	
3016416	4	3016416CD1	25	3016416CB1	7291179CA2
2133755	5	2133755CD1	26	2133755CB1	90067312CA2
5259957	6	5259957CD1	27	5259957CB1	5259957CA2
55029783	7	55029783CD1	28	55029783CB1	4949559CA2
8032202	8	8032202CD1	29	8032202CB1	
6937367	9	6937367CD1	30	6937367CB1	
3876510	10	3876510CD1	31	3876510CB1	
4900076	11	4900076CD1	32	4900076CB1	
1543848	12	1543848CD1	33	1543848CB1	
6254070	13	6254070CD1	34	6254070CB1	767737CA2, 90072067CA2, 9007218CA2
1289839	14	1289839CD1	35	1289839CB1	8698064CA2
5565648	15	5565648CD1	36	5565648CB1	
2764456	16	2764456CD1	37	2764456CB1	
5734806	17	5734806CD1	38	5734806CB1	
7495168	18	7495168CD1	39	7495168CB1	
7483131	19	7483131CD1	40	7483131CB1	
4558650	20	4558650CD1	41	4558650CB1	
7506195	21	7506195CD1	42	7506195CB1	

WO 02/097032

PCT/US02/11152

Table 2

Polypeptide SEQ ID NO:	Accession Polypeptide ID	GenBank ID NO. or PROTEOME ID NO.	Probability Score	Annotation
1	1351608CD1	g3170887	0.0	[Mus musculus] ubiquitin-protein ligase E3-alpha (Kwon, Y.T., et al. (1998) The mouse and human genes encoding the recognition component of the N-end rule pathway) Proc. Natl. Acad. Sci. U.S.A. 95: 7898-7903.)
2	4259314CD1	g11990422	8.2E-198	[Homo sapiens] MOP-4
3	3660046CD1	g12656270	0.0	[Mus musculus] ubiquitin-protein ligase Nedd4-2. Kanyama, E. et al. (2001) A novel mouse Nedd4 protein suppresses the activity of the epithelial Na+ channel. FASEB J. 15:204-214.
4	3016416CD1	g4868435	3.1E-45	[Homo sapiens] apoptosis related protein APR-2
5	2133755CD1	g12248402	8.4E-281	[Homo sapiens] cisplatin resistance related protein CRR9p. Yamamoto, K. et al. (2001) Biochem. Biophys. Res. Commun. 280:1148-1154.
6	5259957CD1	g1903384	3.7E-120	[Homo sapiens] preferentially expressed antigen of melanoma. Ikeda, H., et al. (1997) Characterization of an antigen that is recognized on a melanoma showing partial HLA loss by CTL expressing an NK inhibitory receptor. Immunity 6, 199-208.
7	55029783CD1	g5131896	1.7E-188	[Homo sapiens] bladder cancer overexpressed protein
8	8032202CD1	g9081879	4.3E-11	[Homo sapiens] nasopharyngeal carcinoma-related protein. He, Z.W., et al. (2000) Cloning of a novel gene associated with human nasopharyngeal carcinoma. Chin. Sci. Bull. 45, 2267-2272.
9	6937367CD1	g8650435	1.5E-299	[Mus musculus] RafGDS-like protein 3. Shao, H. and Andres, D.A. (2000) a novel RafGEF-like protein, RGL3, as a candidate effector for raf and Ras. J. Biol. Chem. 275, 26914-26924.
10	3876510CD1	g5870832	7.3E-259	[Mus musculus] skin-BOP1. Hwang, I. and Gottlieb, P.D. (1997) The Bop gene adjacent to the mouse CD86 gene encodes distinct zinc-finger proteins expressed in CTLs and in muscle. J. Immunol. 158, 1165-1174.

WO 02/097032

PCT/US02/11152

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
11	4900076CD1	g5992657	2.2E-158	[Homo sapiens] PR-domain containing protein 9. Jiang, G.L., and Huang, S. (2000) The yin-yang of PR-domain family genes in tumorigenesis. <i>Histol. Histopathol.</i> 15, 109-117.
12	1543848CD1	g4809026	3.2E-121	[Homo sapiens] suppressor of 62 allele of <i>skp1</i> homolog. Sg1. Kitagawa, K., et al. (1999) <i>Mol. Cell</i> 4, 21-33.
13	6254070CD1	g348918	2.5E-205	[Homo sapiens] hWNT5A. Clunk, C. C. et al. (1993) Molecular cloning of the human proto-oncogene Wnt-5A and mapping of the gene (WNT5A) to chromosome 3p14-p21. <i>Genomics</i> 18:249-260.
14	1289839CD1	g1666073	1.7E-48	[Homo sapiens] RRP22 protein. Zucana-Rossi, J. et al. (1996) Identification of new members of the Gsc2 and Ras families in the 22q12 chromosome region. <i>Genomics</i> 38:247-254.
15	5565648CD1	g6164747	5.3E-36	[Homo sapiens] F-box protein Fbx22. (Cenciarelli, C. et al. (1999) Identification of a family of human F-box proteins. <i>Curr. Biol.</i> 9: 1177-1179.)
16	2764456CD1	g6630609	3.6E-252	[Homo sapiens] cyclin-E binding protein 1. (Mitsui, K. et al. (1999) A novel human gene encoding HECT domain and RCC1-like repeats interacts with cyclins and is potentially regulated by the tumor suppressor proteins. <i>Biochem. Biophys. Res. Commun.</i> 266:115-122.)
17	5734806CD1	g5923891	0.0	[Homo sapiens] cyclophilin-related protein Anderson, S.K. et al. (1993) <i>Proc. Natl. Acad. Sci. U.S.A.</i> 90:542-546.
18	7495168CD1	g13277570	6.0E-108	[Homo sapiens] (BC005694) tumor necrosis factor, alpha-induced protein 1 (endolethal)

WO 02/097032

PCT/US02/11152

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
19	748313CD1	g2330015	8.3E-21	[Homo sapiens] apoptotic protease activating factor 1 (Apaf-1) Zou, H. et al. (1997) Cell 90, 405-413
20	4558650CD1	g4416181	2.4E-234	[Mus musculus] F518 Park, E.J. et al. (1999) Nucleic Acids Res. 27, 1524-1530
21	7506195CD1	g4416181	5.2E-155	[Mus musculus] F518 Park, E.J. et al. (1999) <i>supra</i>

WO 02/097032

PCT/US02/11152

Table 3

SEQ ID NO.	Protein ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	1351608CD1	1738	S16 S100 S277 S570 S566 S625 S689 S715 S816 S833 S989 S993 S1053 S1078 S1257 S1272 S1278 S1489 S1582 S1605 T61 T162 T194 T250 T433 T480 T727 T752 T757 T787 T798 T882 T973 T1036 T1080 T1512 T1549 Y99 Y216 Y677	N281 N711 N750 N785 N1043	Puative zinc finger in N-recogin (a recognition component of the N-end rule pathway): Q86-E156	HMMER, PFAM
					Transmembrane domains: V633-E651, F864-R881, S1278-L1295, I1378-Y1406, S1421-L1449, G1483-Y1507, G1640-R1660	TMAP
					UBIQUITIN PROTEIN LIGASE E3 COMPONENT NRECOGNIN UBIQUITIN PROTEIN LIGASE E3ALPHA LIGASE PDI49323: L1202-Y1485	BLAST_PRODOM
					LIGASE UBIQUITIN PROTEIN E3 COMPONENT NENDRECOGNIZING N RECOGNIN UBIQUITIN CONJUGATION PD066492: A355-E897, T752-I1225, I341-I535, F24-S100	BLAST_PRODOM

WO 02/097032

PCT/US02/11152

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	cont				UBIQUITINPROTEIN LIGASE E3 COMPONENT NRECOGNIN UBIQUITINPROTEIN LIGASE E3ALPHA LIGASE PDI49312: M1-F85	BLAST_PRODOM
					LIGASE UBIQUITINPROTEIN E3 COMPONENT NRECOGNIN UBIQUITINPROTEIN N RECOGNIN UBIQUITIN CONJUGATION PD010181: F37-Q284, F37-L229	BLAST_PRODOM
					Leucine zipper pattern: L593-L614	MOTIFS
2	4259314CD1	389	S108 S149 S272 S317 S340 S360 T25 T89 T93 T208 T209 T230 T246 T277 T293 T364 Y132	N143 N220 N257	TH1F family: A60-G200	HMMER_PRAM
					Transmembrane domain: G70-K86 N terminus is non-cytosolic.	TMAP
					UBIQUITINACTIVATING ENZYME E1 UBIQUITIN CONJUGATION LIGASE REPEAT MULTIGENE FAMILY PROTEIN PD005434: C197-E363	BLAST_PRODOM

WO 02/097032

PCT/US02/11152

Table 3

SEQ ID NO.	Protein	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2	cont				PROTEIN ENZYME UBIQUITINACTIVATING BIOSYNTHESIS MULTIGENE PD000731: S62-F196	BLAST_PRODOM
					UBIQUITIN-ACTIVATING ENZYME DM00412P2231444-293: A35-Q290 DM00412P209737-288: E38-I287 DM00412P1252140-291: S33-I287 DM00412P2231518-256: T34-I287	BLAST_DOMO
3	3660046CD1	854	S16 S104 S110 S130 S182 S191 S200 S214 S221 S268 S366 S510 S565 S638 S771 S832 T6 T154 T248 T402 T403 T437 T453 T505 T624 T648 T679 T784 Y688	N59 N95 N287 N569	HECT domain (ubiquitin-transferase): F548-V853	HMMER_Pfam
					WW domains: L266-P295, L378-P407, L74-P103, L429-P458	HMMER_Pfam
					Transmembrane domain: F589-F610	TMAP

WO 02/097032

PCT/US02/11152

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	cont				WW/psf5WWP domain proteins: BL01159: Y89-P103 WW domain signature: PR00403: L378-R391, Y281-P295 HECT-domain (ubiquitin-transferase): PF00632: F711-D717, W761-P788, L815-N846 PROTEIN LIGASE UBIQUITIN CONJUGATION REPEAT UBIQUITIN-PROTEIN DNA BINDING PROBABLE ONCOGENIC PD002225: I623-F850, A543-L669	BLIMPS_BLOCKS BLIMPS_PRINTS BLIMPS_PFAM BLAST_PRODOM
					NEDD4 PROTEIN PD092083: S104-G265 PD064436: T296-F377 PD033209: M1-S63	BLAST_PRODOM
					HECT DOMAIN DM01690P239940513-808: N556-F850 DM01690P25159319-306: N556-F850 DM01690A38919785-1082: F555-D851 DM01690P23119615-909: N556-D851	BLAST_DOMO
					WW/psf5WWP domain signature: W78-P103, W270-P295, W382-P407, W433-P458 signal_cleavage: M1-C26	MOTIFS SPSCAN
4	3016416CD1	111	S69 T88			

WO 02/097032

PCT/US02/11152

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	2133755CD1	538	S3 S34 S161 S219 S231 S268 S317 S322 S387 S389 S491 S526 T60 T160 T385 T470 T476	N91 N101 N229	Signal Peptide: M1-T23 Transmembrane domain: M1-Y28 T282-S310 T323-V349 Y429-K457 P465-R493 N-terminus is cytosolic. PROTEIN THYMIC EPITHELIAL CELL SURFACE ANTIGEN R166.2 T13H5.2 PD025254: E67.P410 PROTEIN R166.2 T13H5.2 PD042934: L411-K530	HMMER TMAP BLAST_PRODOM BLAST_PRODOM
6	5259957CD1	474	S2 S26 S45 S77 S80 S122 S128 S192 S277 S305 S317 T27 T51 T139 T157 T439	N155 N270	Leucine Rich Repeat: Q320-L343, T349-S373, K99-P124 MELANOMA PREFERENTIALLY EXPRESSED ANTIGEN KIAA0014 PROTEIN D085024.3 PRAME LIKE OF PD043129: L69-C472, A5-E125	HMMER, PFAM BLAST_PRODOM
7	55029783CD1	354	S12 S19 S168 S186 S341 T7 T15	N84 N133 N138 N279	Transmembrane domains: T74-L99 S116-S135 L176-V204 Y231-P251 W260-L280 T289-G309 S318-H337 N-terminus is cytosolic	TMAP

WO 02/097032

PCT/US02/11152

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	cont				Cell attachment sequence: R39-D41	MOTIFS
8	8032202CD1	272	S82 S226 T237 T256	N36 N67 N95 N118	Signal Peptide: M2-A20, M2-K23, M1-P25, M1-A32 Transmembrane domain: W6-R34 N-terminus is non-cytosolic	SPSCAN HMMER
9	6937367CD1	710	S30 S36 S40 S170 S221 S247 S277 S342 S377 S388 S402 S495 S512 S520 S526 S540 S559 S580 S591 S601 T63 T99 T256 T290 T372 T522 T635	N339	Ras association (RaGDS/AF-6) domain: E613-G700 RasGAP domain: L244-K454 PROTEIN FACTOR RELEASING GUANINENUCLEOTIDE GUANINE NUCLEOTIDE OF SH3 DOMAIN SON PD002030, L243-W456 LTE1 PROTEIN DM08606/a56234/4-495, L11, S499	HMMER_Pfam BLAST_PRODOM BLAST_DOMO

WO 02/097032

PCT/US02/11152

Table 3

SEQ ID NO.	Incye Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	cont				GUANINE-NUCLEOTIDE DISSOCIATION STIMULATORS CDC25 FAMILY DM01741 P04821 1062-1534; L69-L393, V421-L498	BLAST_DOMO
10	38/6510CD1	490	S128 S243 S251 T183 T444 Y325 Y458	N211 N257	MYND zinc finger: CS2-C90 Transmembrane domain: I348-R376 N-terminus is cytosolic Zinc finger PF00642: L62-H72	HMMER_Pfam TMAP BLMPs_Pfam
					PROTEIN ZINC-FINGER BOP NUCLEAR DNA-BINDING ALTERNATIVE SPLICING METAL-BINDING PD034977: E327-Q490 PD129736: E7-V51 PD129734: Q280-H326	BLAST_PRODOM
					PROTEIN ZINC-FINGER BY EVIDENCE=PREDICTED MATCH PUTATIVE BOPLIKE ORF YEL165C ZINC PD005436: A92-L287	BLAST_PRODOM

WO 02/097032

PCT/US02/11152

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11	4900076CD1	599	S80 S100 S181 S203 S277 S307 S333 S363 S389 S475 S503 S559 S587 T28 T33 T63 T72 T189 T235 T318 T346 T395 T450 Y325 Y457	N44 N311	KRAB box: V62-S112	HMMER_Pfam
					Zinc finger, C2H2 type: Y437-H459, Y521-H543, Y325-H547, Y297-H519, Y577-H599, Y409-H431, Y549-H571, Y269-H291, Y465-H487, Y381-H403, Y495-H515, Y353-H575	HMMER_Pfam
					Zinc finger, C2H2 type BL00028: C299-H315	BLIMPS_PRINTS
					PROTEIN ZINC FINGER ZINC PD01066: P64-A102	BLIMPS_PRODOM
					ZINC FINGER PROTEIN PD017719: P352-F586 PD101192: A102-N242 PD000072: K295-C358	BLAST_PRODOM
					ZINC FINGER, C2H2 TYPE, DOMAIN DM00002(P08042)314-358: C358-H403, C326-H571, C442-H487	BLAST_DOMO

WO 02/097032

PCT/US02/11152

Table 3

SEQ ID NO.	Incye Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11	cont				Zinc finger, C2H2 type, domain: C271-H291 C299-H319 C327-H347 C355-H375 C383-H403 C411-H431 C439-H459 C467-H487 C495-H515 C523-H543 C551-H571 C579-H599	MOTIFS
12	1543848CD1	365	S11 S17 S161 S202 S214 S239 S331 T101 T126 T141 T234 T240 T265 Y47	N77 N78 N146 N159 N329 N343	signal_cleavage: M1-A68	SPSCAN
					TTR Domain: A45-N78, S79-T112	HMMER_PFIAM
					PROTEIN D1054.3 SGT1 PD022846: K194-Y365	BLAST_PRODOM
13	6254070CD1	365	S7 S101 S195 S233 T280 T337 T356	N89 N105 N297 N311	signal_cleavage: M1-A22	SPSCAN
					Signal Peptide: M6-V25, M1-S30, M6-S30	HMMER
					wnt family: L53-K365	HMMER_PFIAM
					Transmembrane domain: S58-W32 N-terminus is non-cytosolic	TMAP
					Wnt-1 family proteins: BL00246: G81-C100, G116-S150, W162-E186, M203-Y255, G319-C364	BLIMPS_BLOCKS
					Wnt-1 family signature wnt.lpf: M203-K252	PROFILESAN
					PROTEIN DEVELOPMENTAL GLYCOPROTEIN PRECURSOR SIGNAL: WNT1 WNT5A WNT2	BLAST_PRODOM
					EXTRACELLULAR MATRIX PD000810: D161-K365, L53-G267	
					WNT-1 FAMILY DM000403:Q66421:13-358: Q23-K365	BLAST_DOMO

WO 02/097032

PCT/US02/11152

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13	cont				WNT-1 FAMILY DM00403 p09544 26-348: L53-K365	BLAST_DOMO
					WNT-1 FAMILY DM00403 p47793 24-351: Q41-K365	BLAST_DOMO
					WNT-1 FAMILY DM00403 p44889 25-359: S30-K365	BLAST_DOMO
14	1289439CD1	203	S159 T4 T37 T70 T105 T150	N142	Wnt-1 family signature C223-C232 Ras family: R6-M203 Transmembrane domain: C79-R107, N-terminus is non-cytosolic Transforming protein P21 RAS signature PR00449: S119-L132, Y155-V177, Y5-Y26, V46-V68	MOTIFS HMMER_PPAM TMAP BLIMPS_PRINTS
					RAS TRANSFORMING PROTEIN DM00006 S41960 3-148: R6-E157	BLAST_DOMO
					RAS TRANSFORMING PROTEIN DM00006 E5540 3-148: R6-E157	BLAST_DOMO
					RAS TRANSFORMING PROTEIN DM00006 P4855 8-153: Y5-E157	BLAST_DOMO
					RAS TRANSFORMING PROTEIN DM00006 P11224 1-157: V2-E157	BLAST_DOMO
					ATP/GTP-binding site motif A (P-loop) G11-S18	MOTIFS
15	5565648CD1	403	S13 S113 S128 S162 S263 S401 T62 T127 T242	N198 N257	R-box domain: S19-T67	HMMER_PPAM

WO 02/097032

PCT/US02/11152

Table 3

SEQ ID NO.	Accession ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15	cont				Transmembrane domain: A344-G361 N-terminus is cytosolic.	TMAP
16	2764456CD1	1022	S46 S216 S224 S242 S290 S339 S368 S385 S397 S647 S669 S690 S793 S801 S807 S867 S915 S975 S995 T117 T118 T124 T259 T319 T323 T362 T375 T398 T427 T431 T443 T448 T459 T550 T552 T729 T850 T904 T939 T951 T973 Y758	N263 N865	HECT-domain (ubiquitin-transferase): E726-V1015	HMMER_PFAM
					Regulator of chromosome condensation (RCC1): K93-K145, D254-H394, D146-L198, G200-Q253, R43-H92	HMMER_PFAM
					Transmembrane domain: R757-K785 N-terminus is non-cytosolic.	TMAP
					Regulator of chromosome condensation (RCC1) proteins: BL00625: G136-K164, Y238-S271	BLIMPS_BLOCKS

WO 02/097032

PCT/US02/11152

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	cont				Regulator of chromosome condensation (RCC1) signatures: rcc1_1.prf: V186-V238 rcc1_2.prf: A60-F12, G110-E165, P221-G269 Chromosome condensation regulator RCC1 signature: PR00633: R95-E111, V149-E165, I241-V257	PROFILES CAN
					HECT-domain (ubiquitin-transferase) PF00632: V836-G842, L930-I957, H979-N1010	BLUMPS, PFAM
					PROTEIN LIGASE UBIQUITIN CONJUGATION REPEAT UBIQUITIN/PROTEIN DNA BINDING PROBABLE ONCOGENIC PD002225: V716-F1014	BLAST_PRODOM
					UBIQUITIN CONJUGATION PD136613: V348-L674, P68-G110, F119-G163, E11-Q70, P227-G307, Q169-L228 PD071828: K104-L198	BLAST_PRODOM
					PROTEIN REPEAT GUANINE-NUCLEOTIDE RELEASING FACTOR REGULATOR CELL CYCLE MITOSIS PD001424: K93-G163	BLAST_PRODOM

WO 02/097032

PCT/US02/11152

Table 3

SEQ ID NO.	Lucyve Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	cont				HECT DOMAIN DM01690[P29940]S13-808: F721-F1014 DM01690[P40985]S78-891: F731-D905, V836-F1014, Q212-K226 DM01690[A389197]S5-1082: Y738-F1014 DM01690[P531196]S-909: K755-F1014 Regulator of chromosome condensation (RCC1) signature 2: A28-L38, V80-V90, V133-L143, V186-L196, I241-L251	BLAST_DOMO
17	S734806CD1	1462	S145 S180 S206 S210 S212 S214 S221 S235 S243 S244 S260 S274 S328 S348 S350 S352 S383 S393 S394 S402 S462 S466 S470 S476 S483 S494 S495 S500 S502 S513 S517 S531 S553 S563 S565 S670 S681 S688 S695 S722 S747 S761 S766 S768 S773 S774 S776 S798 S802	N119 N203 N863 N870 N900 N937 N1047	Cyclophilin-type peptidyl prolyl cis-trans isomerase: Q8-A177	HMMER_FFAM

WO 02/097032

PCT/US02/11152

Table 3

SEQ ID NO.	Protein ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17			S808 S820 S846 S848 S855 S882 S888 S890 S898 S906 S915 S917 S939 S949 S955 S957 S964 S1044 S1060 S1061 S1066 S1076 S1180 S1202 S1243 S1275 S1295 S1296 S1303 S1318 S1346 S1372 S1375 S1378 S1382 S1390 S1398 S1417 S1435 S1441 S1443		Transmembrane domain: P131-I150 N-terminus is non-cytosolic	TMAP
			S1445 S1454 S1458 T35 T37 T126 T193 T337 T419 T453 T471 T839 T931 T989 T1049 T1057 T1082 T1207 T1252 T1283 Y90 Y1461		Cyclophilin-type peptidyl-prolyl cis-trans isomerase signature BL00170: G21-K47, Y59-N98, A106-I150	BLIMPS_BLOCKS
					Cyclophilin-type peptidyl-prolyl cis-trans isomerase signature & profile: D30-D96	PROFILES CAN

WO 02/097032

PCT/US02/11152

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17					Cyclophilin peptidyl-prolyl cis-trans isomerase signature PR00153: L27-L42, P64-G76, F107-Q122, Q122-D134, G135-I150	BLIMPS_PRINTS
cont					NK TUMOR RECOGNITION PROTEIN NATURAL KILLER CELLS CYCLOPHILIN RELATED NKTR CYCLOSPORIN ISOMERASE ROTAMASE REPEAT TRANSMEMBRANE PD145736: N729-T1328 PD078637: G557-Y680 PD065318: C447-G530	BLAST_PRODOM
					ISOMERASE ROTAMASE CYCLOPHILIN CIS- TRANS PEPTIDYL-PROLYL PHASE CYCLOSPORIN MULTIGENE FAMILY PROTEIN PD000341: F11-L176	BLAST_PRODOM
					RECOGNITION; TUMOR; PROLYL; NATURAL; DM08077	BLAST_DOMO
					P230414230-1403: L176-R1348 B447328230-1507: L176-Y1461	
					CYCLOPHILIN-TYPE PEPTIDYL-PROLYL CIS- TRANS ISOMERASE DM00129	BLAST_DOMO
					P23041458-229: Q4-V175	
					Cyclophilin-type peptidyl-prolyl cis-trans isomerase signature: Y59-G76	MOTIFS

WO 02/097032

PCT/US02/11152

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
18	7495168CD1	329	S22 S71 S89 S110 S155 S186 T55 T60 T143 T166T187 T245 T250 T271 T315	N176	K+ channel tetramerisation domain: K41-Q138	HMMER_FFAM
19	7483131CD1	476	S38 S128 S136 S159 S175 S194 S333 S370 S439 S445 T7 T33 T164 T184 T268 T444 T462	N150 N365 N460	PROTEIN EDP1 TUMOR NECROSIS FACTOR ALPHA-INDUCED ENDOTHELIAL B12 PD037429: L118-D329 signal_cleavage: M1-L30	BLAST_PRODOM SPSCAN
					SAM domain (Sterile alpha motif): E330-R394 WD domain, G-beta repeat: L4-S38, I273-Q309, Y231-D267, F46-N82, A90-N125, C172-I218, L131-D167	HMMER_FFAM
					Trp-Asp (WD-40) repeats signature: L243-A290 Beta G-protein (transducin) signature PR00319: L235-G251, L254-T268, I273-A288, P291-W308 G-protein beta WD-40 repeat signature PR00320: L69-T83, L296-F310	PROTELESCAN BLIMPS_PRINTS
20	4558650CD1	485	S119 S170 S343 S363 T155 T312		Trp-Asp (WD) repeats signature: L69-T83, L112-A126 signal_cleavage: M1-A68	MOTIFS SPSCAN

WO 02/097032

PCT/US02/11152

Table 3

SEQ ID NO.	Lucyfe Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
20	cont				PROTEIN REPEAT SIGNAL PRECURSOR PRION GLYCOPROTEIN NUCLEAR GPI ANCHOR BRAIN MAJOR PD001091: G10-P123 PROTEIN TYROSINE PHOSPHATASE TD14 EC 3.1.3.48 HYDROLASE PD180360: P4-R187, E323-K357	BLAST_PRODOM
					FORMIN; DM04565[Q08588]1-1212: P12-G115 FIBRILLAR COLLAGEN CARBOXYL-TERMINAL DM00043[A41132]43-133: P12-C95 DM00042[S21930]37-137: P12-F98 PROLINE-RICH PROTEIN DM03894[A39066]1-159: P14-P129	BLAST_DOMO
21	7506195CD1	406	S119 S170 T153 T312		signal_cleavage: M1-A68 PROTEIN REPEAT SIGNAL PRECURSOR PRION GLYCOPROTEIN NUCLEAR GPI ANCHOR BRAIN MAJOR PD001091: G10-P123, R11-P129	SPSCAN BLAST_PRODOM
					FIBRILLAR COLLAGEN CARBOXYL-TERMINAL DM00043[A41132]43-133: P12-C95, P14-P117	BLAST_DOMO
					FIBRILLAR COLLAGEN CARBOXYL-TERMINAL DM00043[S21930]37-137: P12-F98, P20-P120	BLAST_DOMO
					PROLINE-RICH PROTEIN DM03894[A39066]1-159: P14-P129	BLAST_DOMO
					FORMIN; DM04565[Q08588]1-1212: P12-G115	BLAST_DOMO

WO 02/097032

PCT/US02/11152

Table 4

Polymer SEQ ID NO. / Accession ID / Sequence Length	Sequence Fragments
22/1351608CB1/ 7742	1-644, 1-733, 637-5226, 3489-3638, 3563-3836, 3661-4146, 4346-4436, 4349-4563, 4437-4914, 4545-4994, 4559-4972, 4574-4755, 4635-4889, 4660-5217, 4678-4738, 4678-4816, 4683-4816, 4687-4824, 4689-4998, 4689-5130, 4689-5287, 4801-5354, 4802-4985, 4812-5178, 4818-5380, 4889-5124, 4896-5198, 4904-5425, 4929-5381, 4981-5151, 4981-5249, 4981-5565, 4993-5228, 5005-5380, 5098-5321, 5126-5401, 5127-5434, 5270-5519, 5271-5560, 5324-5605, 5384-5802, 5386-5992, 5431-5703, 5431-5881, 5431-5938, 5431-5944, 5431-6023, 5469-5733, 5499-5939, 5521-5677, 5596-6077, 5630-5831, 5739-6288, 5745-5984, 5748-5999, 5779-6051, 5799-6068, 5860-6312, 5910-6472, 5940-6491, 5954-6435, 5954-6459, 6135-6600, 6139-6600, 6188-6520, 6193-6462, 6227-6470, 6389-6352, 6290-6487, 6319-6582, 6333-6584, 6348-6625, 6401-6679, 6401-6916, 6441-6698, 6441-6802, 6443-6705, 6449-6702, 6470-6727, 6480-6727, 6502-6792, 6534-6869, 6588-6876, 6591-6864, 6591-6875, 6628-6835, 6671-6929, 6707-6880, 6711-7010, 6751-7095, 6778-7063, 6843-7113, 6855-7102, 6880-7134, 6899-7153, 6900-7186, 6947-7228, 6948-7200, 7030-7640, 7050-7638, 7087-7613, 7087-7622, 7091-7650, 7093-7634, 7105-7345, 7151-7631, 7161-7453, 7161-7649, 7161-7665, 7173-7650, 7176-7656, 7186-7659, 7192-7651, 7194-7650, 7199-7650, 7200-7653, 7202-7650, 7204-7650, 7214-7655, 7224-7650, 7228-7650, 7244-7498, 7249-7656, 7259-7650, 7268-7650, 7269-7655, 7269-7680, 7280-7650, 7292-7522, 7311-7650, 7316-7652, 7319-7650, 7322-7655, 7325-7650, 7331-7650, 7341-7640, 7353-7650, 7364-7659, 7380-7650, 7396-7653, 7399-7651, 7403-7647, 7412-7644, 7413-7663, 7414-7647, 7417-7607, 7417-7648, 7417-7661, 7417-7667, 7497-7650, 7502-7657, 7520-7652, 7533-7742, 7535-7656, 7592-7655
23/4259314CB1/ 1674	1-73, 1-121, 1-138, 1-217, 1-337, 1-341, 1-363, 1-418, 1-430, 1-440, 1-463, 1-470, 1-476, 1-484, 1-494, 1-508, 1-513, 1-521, 1-526, 1-628, 1-637, 2-232, 4-93, 4-246, 4-544, 4-588, 4-647, 4-681, 5-593, 5-604, 5-636, 16-526, 16-630, 20-476, 21-694, 77-680, 120-726, 195-778, 277-898, 411-1089, 442-1087, 534-1126, 535-1149, 551-1254, 562-819, 562-1040, 574-1166, 590-881, 591-1264, 670-1352, 673-1154, 677-1126, 696-1318, 721-1394, 726-980, 726-1151, 735-1336, 740-1456, 745-1126, 762-1456, 766-1252, 772-1213, 794-1456, 803-1289, 815-920, 837-1059, 893-1351, 897-1671, 900-1133, 927-1653, 961-1642, 968-1671, 1014-1633, 1023-1126, 1040-1641, 1054-1669, 1054-1674, 1055-1126, 1059-1126, 1059-1383, 1068-1650, 1161-1671, 1167-1671, 1179-1671, 1184-1674, 1191-1671, 1242-1656, 1259-1674

WO 02/097032

PCT/US02/11152

Table 4

Polynucleotide SEQ ID NO./ Inventor ID/ Sequence Length	Sequence Fragments
2473660046CB1/ 3671	1-827, 5-839, 11-599, 96-999, 237-558, 247-757, 348-932, 367-886, 376-2955, 438-700, 438-956, 633-1189, 648-1093, 828-1461, 955-1153, 1404-1591, 1660-2067, 1682-3032, 1709-2307, 1841-2388, 1925-2601, 1937-2463, 1989-2760, 2076-2486, 2167-2759, 2263-3046, 2441-2705, 2574-3040, 2634-2899, 2734-3297, 2736-3248, 2764-3070, 2843-3052, 2853-3220, 2860-3338, 2911-3511, 3007-3655, 3013-3461, 3086-3367, 3094-3325, 3114-3671
257016416CB1/ 2038	1-608, 170-2031, 371-608, 410-1096, 422-868, 429-963, 515-775, 515-1114, 519-1124, 579-926, 600-1266, 623-999, 623-1222, 661-1207, 661-1284, 731-1290, 762-1339, 852-961, 866-961, 947-1534, 963-1165, 1016-1544, 1026-1582, 1034-1339, 1068-1232, 1095-1383, 1273-1550, 1273-1786, 1301-1864, 1327-1540, 1407-1686, 1425-1880, 1445-2021, 1484-1991, 1534-2003, 1581-1894, 1635-1904, 1639-1928, 1682-1992, 1682-1999, 1682-2038, 1697-1960, 1697-1971
262133755CB1/ 2235	1-263, 69-584, 107-676, 107-718, 107-755, 109-300, 109-457, 109-480, 109-554, 109-566, 109-627, 109-657, 109-670, 109-720, 109-786, 109-788, 109-885, 111-200, 111-276, 111-377, 111-518, 111-560, 111-575, 111-577, 111-606, 111-610, 111-652, 111-661, 111-680, 111-697, 111-706, 111-749, 111-770, 111-782, 111-788, 111-840, 112-207, 114-697, 116-841, 117-260, 118-401, 118-576, 119-635, 127-543, 127-592, 128-808, 131-736, 132-371, 132-378, 132-444, 132-541, 132-601, 132-609, 209-400, 222-478, 283-542, 350-574, 365-779, 375-601, 405-608, 405-624, 633-890, 661-905, 697-964, 717-970, 720-939, 720-958, 720-968, 723-1039, 744-1011, 763-1053, 771-1001, 808-1276, 840-1264, 845-1161, 848-1085, 865-1167, 875-1204, 904-1177, 904-1329, 914-1180, 918-1198, 923-1194, 923-1366, 927-1384, 930-1210, 943-1155, 948-1156, 950-1200, 950-1226, 951-1501, 952-1254, 959-1200, 968-1287, 971-1186, 971-1206, 974-1200, 984-1255, 984-1553, 994-1276, 997-1264, 1023-1294, 1030-1232, 1034-1261, 1036-1263, 1036-1316, 1042-1250, 1046-1319, 1047-1313, 1051-1282, 1056-1327, 1057-1276, 1060-1340, 1064-1317, 1077-1328, 1082-1355, 1085-1354, 1127-1372, 1128-1563, 1135-1398, 1145-1402, 1145-1413, 1156-1458, 1167-1461, 1168-1333, 1180-1389, 1189-1438, 1191-1478, 1194-1509, 1225-1455, 1225-1646, 1227-1487, 1245-1567, 1260-1508, 1261-1594, 1269-1515, 1271-1506, 1273-1411, 1273-1516, 1277-1537, 1302-1483, 1307-1883, 1337-1556, 1338-1538, 1338-1544, 1350-1608, 1350-1619, 1370-1635, 1377-1692, 1384-1641, 1390-1676, 1393-1646, 1394-1623, 1399-1646, 1400-1635, 1409-1678, 1418-1721, 1419-1698, 1450-1708, 1458-1674, 1439-1673, 1439-1960, 1443-1706, 1460-1725, 1460-1983, 1528-1718, 1557-2206, 1564-1815, 1565-1762, 1607-1811, 1611-1840, 1627-2201, 1633-1816, 1643-2214, 1654-1914, 1704-2220, 1803-2026, 1820-2048, 1820-2055, 1850-2120, 1850-2220, 1890-2108, 1893-2113, 1893-2133, 2003-2206, 2048-2219, 2053-2235

WO 02/097032

PCT/US02/11152

Table 4

Polymer SEQ ID NO./ Incr ID / Sequence Length	Sequence Fragments
27/6259957CB1/ 1851	1-346, 1-461, 1-478, 1-526, 1-569, 1-570, 53-550, 248-517, 259-709, 292-913, 312-820, 317-688, 346-919, 360-1014, 361-926, 455-1063, 510-1126, 526-1093, 539-1080, 625-1201, 630-1263, 632-1215, 690-1182, 696-1164, 696-1319, 708-1303, 785-1219, 885-1468, 888-1245, 898-1354, 905-1083, 915-1171, 966-1442, 1074-1673, 1108-1701, 1210-1851, 1226-1531, 1232-1743, 1300-1830
28/55029783CB1/ 1466	1-614, 110-817, 112-683, 117-675, 125-633, 126-697, 141-701, 141-765, 141-847, 141-851, 146-535, 149-700, 155-752, 155-766, 155-783, 155-784, 164-827, 165-789, 168-835, 168-855, 168-862, 170-644, 173-194, 207-779, 215-748, 236-434, 272-743, 405-608, 429-1116, 454-710, 454-938, 467-1116, 471-718, 487-1277, 500-734, 503-1116, 519-1132, 560-745, 560-1369, 561-1305, 564-1132, 638-1369, 683-1342, 977-1211, 977-1466, 1066-1249, 1081-1298
29/8032202CB1/ 1049	1-312, 1-655, 3-834, 256-974, 343-593, 412-1049, 439-731, 439-909, 539-963
30/6937367CB1/ 2520	1-474, 3-121, 3-712, 44-314, 91-198, 91-206, 94-440, 175-600, 207-413, 210-675, 376-1202, 456-1090, 474-958, 524-1054, 573-1119, 554-746, 596-1190, 613-1202, 646-1054, 726-1108, 750-871, 760-1281, 783-903, 808-1056, 849-1300, 878-1441, 919-1678, 985-1242, 1027-1388, 1128-1413, 1144-1688, 1158-1279, 1243-1512, 1243-1576, 1252-1449, 1268-1769, 1269-1581, 1331-1593, 1348-1418, 1351-1388, 1386-1421, 1408-1995, 1419-1463, 1419-1579, 1419-1631, 1419-1689, 1419-1707, 1419-1755, 1419-1789, 1419-1822, 1419-1866, 1419-1872, 1419-1971, 1419-2020, 1419-2058, 1436-1598, 1446-1637, 1452-1913, 1474-1664, 1474-1666, 1480-1943, 1482-1692, 1510-1959, 1543-1681, 1546-2072, 1585-2284, 1604-2123, 1604-2252, 1610-1762, 1617-1887, 1658-1926, 1716-2018, 1752-1951, 1752-2229, 1753-2027, 1770-2054, 1822-2234, 1835-2109, 1855-1996, 1855-2336, 1890-2187, 1933-2137, 1934-2487, 1979-2253, 2019-2520, 2073-2312, 2092-2296, 2120-2347

WO 02/097032

PCT/US02/11152

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments	
3173876510CB1/ 4360	1-619, 1-634, 6-322, 12-786, 14-140, 92-787, 147-463, 147-718, 147-865, 147-879, 147-917, 147-920, 147-927, 147-931, 150-752, 150-860, 150-884, 164-473, 195-1051, 315-429, 315-532, 325-1145, 454-1042, 454-1045, 592-1493, 610-1485, 630-757, 700-1454, 701-913, 744-1032, 767-1457, 770-1457, 804-1126, 923-1154, 923-1375, 942-1197, 957-1239, 960-1241, 1114-1792, 1155-1179, 1164-1441, 1165-1804, 1171-1198, 1236-1703, 1249-1271, 1254-1588, 1321-1580, 1369-1865, 1456-1818, 1457-1751, 1466-1819, 1573-1734, 1727-1983, 1756-2047, 1813-2114, 1839-2115, 1858-2443, 1862-2154, 1863-2371, 1863-2384, 1863-2498, 1863-2498, 1886-2165, 1924-2426, 2023-2363, 2114-2394, 2114-2693, 2114-2803, 2114-2897, 2135-2663, 2176-2726, 2192-2473, 2200-2402, 2201-2438, 2235-2504, 2250-2832, 2352-2696, 2364-2598, 2364-2885, 2364-2910, 2364-2911, 2364-2976, 2364-2988, 2364-3032, 2364-3068, 2364-3069, 2364-3172, 2388-2660, 2389-2749, 2392-2749, 2397-3276, 2405-2696, 2444-3041, 2456-3159, 2463-2699, 2482-3112, 2490-3076, 2492-2732, 2516-3037, 2544-3174, 2557-3214, 2560-3109, 2562-2833, 2570-3277, 2588-3281, 2603-2816, 2606-2886, 2637-3118, 2640-3150, 2655-3152, 2676-3120, 2679-3362, 2694-3217, 2724-3161, 2728-3165, 2739-3205, 2744-3390, 2746-3418, 2756-3401, 2762-3217, 2762-3391, 2766-3182, 2768-3391, 2782-3398, 2801-3284, 2810-3007, 2810-3311, 2819-3289, 2856-3425, 2917-3425, 2942-3494, 2992-3445, 3005-3449, 3020-3672, 3029-3852, 3032-3652, 3040-3545, 3051-3594, 3063-3568, 3065-3553, 3080-3683, 3099-3612, 3113-3792, 3124-3613, 3152-3363, 3156-3626, 3194-3529, 3198-3466, 3213-3237, 3228-3792, 3252-3559, 3257-3971, 3266-3566, 3279-3535, 3296-3990, 3322-3849, 3335-3829, 3338-3956, 3349-3605, 3349-3793, 3349-3910, 3387-3889, 3404-4215, 3441-4333, 3455-4149, 3465-4055, 3479-4071, 3541-4037, 3553-4209, 3566-4133, 3589-3831, 3594-4209, 3646-4353, 3648-4223, 3649-4146, 3661-4161, 3687-3904, 3692-3944, 3693-4360, 3700-4104, 3706-3950, 3706-4157, 3706-4216, 3706-4250, 3706-4339, 3706-4355, 3723-3947, 3723-3950, 3734-3967, 3744-4323, 3749-4190, 3779-4273, 3794-4088, 3815-4346, 3855-4338, 3863-4346, 3868-4183, 3881-4357, 3891-4333, 3901-3936, 3907-4346, 3916-4345, 3918-4303, 3918-4358, 3921-4360, 3924-4346, 3933-4178, 3938-4358, 3953-4349, 3965-4347, 3986-4343, 3988-4360, 3998-4345, 4024-4343, 4035-4239, 4086-4354, 4133-4344, 4153-4342	

WO 02/097032

PCT/US02/11152

Table 4

Polynucleotide SEQ ID NO./ Nucleic Acid Sequence	Sequence Fragments
3500	1-239, 1-242, 1-286, 101-246, 102-359, 105-273, 105-347, 105-407, 108-457, 196-312, 223-708, 319-519, 456-968, 516-642, 643-954, 643-1023, 643-1119, 643-1281, 651-1281, 673-1281, 854-1325, 854-1338, 854-1341, 854-1386, 854-1508, 857-1323, 859-1555, 916-1534, 1035-1636, 1040-1618, 1059-1335, 1061-1276, 1065-1624, 1107-1519, 1114-1201, 1114-1237, 1114-1250, 1114-1392, 1114-1459, 1114-1463, 1114-1464, 1114-1477, 1114-1481, 1116- 1759, 1123-1632, 1135-1235, 1135-1237, 1135-1241, 1135-1248, 1135-1272, 1135-1290, 1135-1300, 1135-1308, 1135-1328, 1135-1340, 1135-1342, 1135-1370, 1135-1380, 1135-1403, 1135-1405, 1135-1423, 1135-1435, 1135- 1436, 1135-1456, 1135-1464, 1135-1508, 1135-1546, 1135-1548, 1135-1560, 1135-1627, 1135-1631, 1135-1632, 1135-1666, 1135-1715, 1147-1201, 1148-1405, 1149-1592, 1158-1582, 1158-1659, 1160-1332, 1182-1465, 1182-1632, 1187-1473, 1329, 1158-1336, 1158-1562, 1158-1575, 1158-1582, 1211-1477, 1217-1287, 1217-1288, 1217-1316, 1217-1323, 1217-1424, 1188-1477, 1188-1716, 1211-1464, 1211-1477, 1217-1287, 1217-1288, 1217-1316, 1217-1323, 1217-1424, 1217-1474, 1217-1540, 1217-1558, 1217-1575, 1217-1703, 1217-1728, 1217-1730, 1218-1405, 1218-1459, 1218- 1477, 1218-1703, 1218-1772, 1219-1359, 1219-1405, 1219-1409, 1219-1416, 1219-1425, 1219-1465, 1219-1558, 1219-1592, 1219-1676, 1219-1722, 1233-1575, 1233-1591, 1235-1575, 1239-1733, 1240-1827, 1245-1477, 1249- 1538, 1251-1575, 1263-1716, 1265-1883, 1266-1812, 1271-1416, 1271-1676, 1273-1518, 1281-1632, 1297-1772, 1298-1880, 1302-1541, 1302-1575, 1302-1735, 1302-1764, 1302-1812, 1303-1642, 1303-1687, 1303-1695, 1303- 1772, 1303-1800, 1317-1672, 1319-1632, 1321-1575, 1323-1827, 1333-1918, 1347-1810, 1349-1967, 1356-1592, 1362-1592, 1362-1632, 1364-1884, 1368-1716, 1377-1896, 1383-1632, 1385-1832, 1386-1632, 1386-1876, 1387- 1487, 1387-1522, 1387-1528, 1387-1575, 1387-1577, 1387-1584, 1387-1593, 1387-1594, 1387-1738, 1387-1759, 1387-1760, 1387-1870, 1387-1876, 1387-1884, 1387-2015, 1391-1968, 1401-1716, 1403-1716, 1404-1911, 1410- 1632, 1417-1708, 1421-1718, 1422-1716, 1431-1884, 1441-2024, 1446-1676, 1446-1982, 1449-1810, 1461-1980, 1469-1716, 1469-1948, 1470-1716, 1470-1760, 1490-1716, 1490-1810, 1491-1984, 1492-2062,

WO 02/097032

PCT/US02/11152

Table 4

Polynucleotide SEQ ID NO./ Insite ID/ Sequence Length	Sequence Fragments
32	1494-1600, 1494-1659, 1494-1661, 1494-1668, 1494-1676, 1494-1810, 1494-1832, 1494-1928, 1494-1968, 1494- 2012, 1495-1832, 1501-1928, 1509-1982, 1515-1735, 1515-1810, 1521-1668, 1523-2135, 1532-1760, 1532-2064, 1533-1884, 1534-1672, 1534-1695, 1534-1738, 1534-1750, 1534-1761, 1534-1885, 1534- 1927, 1534-1939, 1534-1947, 1534-2024, 1534-2052, 1538-2096, 1545-2064, 1549-2024, 1553-1764, 1553-1810, 1554-1810, 1554-2014, 1559-2239, 1569-1884, 1571-1884, 1573-1810, 1575-1982, 1575-2079, 1585-1874, 1590- 1884, 1608-2062, 1608-2146, 1609-1885, 1616-1844, 1616-2131, 1617-1982, 1633-2155, 1637-1884, 1638-1884, 1638-2131, 1638-2146, 1639-1762, 1639-1827, 1639-1829, 1639-1836, 1639-1846, 1639-2002, 1639-2012, 1639- 2122, 1639-2128, 1639-2136, 1643-2146, 1651-2011, 1652-2162, 1653-1982, 1655-1982, 1657-1884, 1659-2062, 1659-2136, 1677-2294, 1683-1947, 1683-1969, 1683-1982, 1690-2294, 1691-2136, 1708-1928, 1708-2146, 1713- 2150, 1722-1940, 1722-1982, 1722-2012, 1722-2098, 1742-2294, 1743-2146, 1744-2150, 1746-1982, 1782-2291, 1785-2012, 1785-2062, 1785-2136, 1785-2146, 1786-1924, 1786-1947, 1786-1993, 1786-1997, 1786-2002, 1786-2013, 1786-2014, 1786-2136, 1786-2149, 1786-2154, 1788-2123, 1797-2136, 1801-2123, 1805-1863, 1805-1869, 1805-1911, 1805-1912, 1805- 1920, 1805-1928, 1805-2014, 1805-2062, 1805-2095, 1805-2107, 1805-2150, 1805-2162, 1806-2062, 1806-2150, 1818-2294, 1823-2294, 1824-2136, 1825-2062, 1827-2136, 1837-2126, 1839-2294, 1851-2150, 1853-2162, 1860- 2002, 1861-2136, 1866-2096, 1866-2136, 1869-1948, 1869-2146, 1888-2136, 1889-2294, 1890-2136, 1890-2137, 1890-2146, 1890-2162, 1890-2353, 1890-2433, 1891-1948, 1891-2020, 1891-2024, 1891-2079, 1891-2081, 1891- 2088, 1891-2097, 1891-2131, 1891-2135, 1891-2136, 1891-2160, 1905-2146, 1907-2146, 1909-2294, 1913-2150, 1914-2136, 1914-2146, 1920-2150, 1929-2136, 1935-2146, 1935-2162, 1956-2136, 1957-2149, 1957- 2150, 1961-2149, 1965-2136, 1974-2088, 1974-2146, 1974-2160, 1974-2162, 1989-2150, 1991-2150, 1994-2136, 1995-2162, 1996-2136, 1998-2146, 2034-2136, 2037-2146, 2037-2150, 2037-2162, 2038-2125, 2040-2149, 2049- 2136, 2053-2136, 2057-2150, 2058-2136, 2058-2150, 2076-2136, 2077-2150, 2079-2146, 2079-2149, 2094-2136, 2113-2149, 2256-2450, 2269-2529, 2269-2859, 2293-2585, 2372-2423, 2372-2429, 2389-2706, 2391-2689, 2409-2668, 2463-2708, 2465-2715, 2518-2814, 2567-3163, 2727-3405, 2752-3404, 2766-3401, 2815- 3067, 2818-3452, 2901-3133, 2901-3167, 2962-3457, 3003-3248, 3103-3487, 3276-3500

Table 4

[illegible]

WO 02/097032

PCT/US02/11152

Table 4

Polynucleotide SEQ ID NO./ Incid ID/ Sequence Length	Sequence Fragments
35/1289839CB1/ 1157	1-699, 47-616, 198-694, 372-1157, 679-927, 918-1157
36/5565648CB1/ 1418	1-297, 1-510, 1-1380, 4-210, 7-693, 16-681, 24-251, 33-293, 41-519, 42-610, 42-674, 48-224, 59-257, 59-613, 71-350, 71-597, 72-399, 75-288, 77-385, 83-341, 84-373, 86-443, 86-492, 86-637, 91-384, 91-640, 91-660, 94-238, 94-345, 94-365, 94-580, 94-632, 96-517, 97-210, 97-395, 101-298, 101-299, 101-362, 102-341, 233-378, 270-540, 378-764, 450-741, 459-668, 463-662, 463-1028, 484-759, 518-764, 531-781, 532-825, 609-844, 634-891, 654-869, 665-890, 680-1127, 686-987, 686-992, 686-1028, 688-1323, 700-1322, 727-1375, 781-1418, 784-1398, 792-1338, 844-1529, 892-1333, 893-917, 893-973, 893-1111, 893-1191, 893-1339, 893-1418, 896-1329, 925-1204, 925-1219, 926-1172, 962-1226, 984-1233, 1127-1327, 1179-1336
37/264456CB1/ 4113	1-195, 37-597, 153-380, 168-597, 224-597, 239-834, 257-968, 305-546, 404-647, 539-759, 559-890, 541-1330, 610-834, 741-1269, 741-1365, 788-977, 1000-1470, 1000-1546, 1000-1674, 1138-1410, 1332-1956, 1343-1506, 1343-1508, 1405-1956, 1441-1773, 1459-1956, 1680-1929, 1686-2024, 1720-1988, 1722-2086, 1783-2262, 1818-2052, 2053-2632, 2059-2360, 2134-2591, 2134-2742, 2174-2472, 2174-2482, 2174-2607, 2175-2564, 2191-2822, 2196-2773, 2205-2500, 2238-2825, 2250-2600, 2324-2599, 2341-2779, 2402-3009, 2405-2471, 2450-2818, 2453-2908, 2442-2699, 2442-2893, 2453-2642, 2467-2516, 2505-2926, 2520-2765, 2520-2780, 2555-3147, 2560-3095, 2565-2840, 2569-3146, 2572-3146, 2588-2878, 2598-3194, 2676-3281, 2687-3164, 2717-3111, 2726-3376, 2780-3189, 2802-3111, 2859-3103, 2876-3569, 2890-3150, 2907-3454, 2920-3188, 2927-3171, 2959-3485, 3114-3687, 3116-3635, 3156-3535, 3156-3726, 3181-3478, 3191-3795, 3192-3487, 3198-3364, 3196-3449, 3199-3772, 3208-3795, 3214-3848, 3219-3721, 3223-3492, 3241-3475, 3241-3529, 3241-3814, 3242-3757, 3246-3476, 3246-3476

WO 02/097032

PCT/US02/11152

Table 4

Polynucleotide SEQ ID NO./ Inocyte ID/ Sequence Length	Sequence Fragments
	3250-3755, 3252-3536, 3252-3744, 3299-3550, 3299-3859, 3340-3877, 3342-3578, 3351-4083, 3357-3592, 3361-3751, 3380-4072, 3393-3688, 3401-3610, 3438-3915, 3453-4055, 3469-3704, 3472-4090, 3476-3847, 3487-4096, 3493-3708, 3515-4085, 3517-4094, 3524-4084, 3542-4064, 3545-3803, 3545-4103, 3551-3813, 3570-3814, 3570-4113, 3571-4092, 3585-3850, 3590-4102, 3593-3834, 3593-4063, 3601-4072, 3606-4110, 3620-3859, 3621-4112, 3623-3946, 3624-4101, 3626-3926, 3641-4108, 3648-4103, 3650-3890, 3656-4103, 3658-3947, 3660-4053, 3666-4113, 3669-4102, 3673-3849, 3678-3932, 3680-4103, 3693-4113, 3696-3920, 3705-4023, 3705-4103, 3711-3790, 3712-3762, 3712-3778, 3712-3790, 3712-3790, 3716-4104, 3719-3790, 3723-4104, 3727-3779, 3727-3788, 3732-4113, 3735-4107, 3739-4112, 3742-3790, 3748-3790, 3749-4102, 3787-4103, 3792-4105, 3803-3869, 3803-3881, 3810-3869, 3810-4021, 3818-3869, 3818-4100, 3823-3869, 3830-3879, 3833-4102, 3838-4113, 3850-4106, 3857-4113, 3874-4102, 3895-4103, 3977-4111
385734806CB1/ 7058	1-789, 1-7058, 454-485, 488-716, 488-959, 488-1017, 491-1035, 524-753, 713-1374, 1025-1321, 1236-1847, 1268-1516, 1284-1543, 1290-1695, 1302-1585, 1310-1860, 1318-2019, 1329-1557, 1350-1952, 1383-1561, 1390-1643, 1392-1943, 1403-1664, 1432-1804, 1483-1776, 1484-1789, 1484-1800, 1488-2101, 1489-1775, 1521-1777, 1562-1815, 1584-1881, 1597-1882, 1614-2220, 1636-2100, 1676-2074, 1697-1989, 1740-2046, 1741-2004, 1759-2059, 1768-2045, 1774-2079, 1832-2102, 1840-2386, 1845-2120, 1847-2088, 1853-2099, 1857-2138, 1888-2370, 2009-2277, 2040-2313, 2090-2586, 2092-2352, 2104-2386, 2108-2327, 2181-2434, 2200-2804, 2265-2962, 2270-2915, 2282-2926, 2294-2516, 2329-2595, 2332-2612, 2355-2582, 2359-2614, 2360-3027, 2445-2722, 2484-3041, 2525-3158, 2543-3133, 2636-2940, 2642-3238, 2647-3103, 2651-2932, 2678-2885, 2729-3077, 2737-3015, 2737-3117, 2739-3039, 2739-3103, 2740-3310, 2754-3368, 2807-3148, 2850-3308, 2876-3486, 2992-3498, 3079-3367, 3113-3665, 3119-3628, 3154-3806, 3183-3747, 3265-3880, 3272-3915, 3323-3949, 3410-3726, 3427-3898,

WO 02/097032

PCT/US02/11152

Table 4

Polynucleotide SEQ ID NO./ Nucleic Acid Sequence Length	Sequence Fragments
	3447-3874, 3448-3958, 3493-3781, 3493-3938, 3539-4077, 3597-4291, 3644-3937, 3694-3874, 3694-3903, 3750-3901, 3799-3972, 3970-4531, 4007-4268, 4073-4663, 4122-4365, 4139-4435, 4152-4607, 4260-4521, 4260-4754, 4368-4675, 4368-4829, 4400-4676, 4408-4628, 4411-4725, 4434-4875, 4522-4752, 4537-4772, 4539-4798, 4539-4994, 4565-4756, 4651-4889, 4653-4901, 4681-4935, 4717-4966, 4717-5203, 4717-5240, 4772-5022, 4802-5092, 4804-5073, 4813-5061, 4830-5176, 4831-5115, 4832-5084, 4908-5328, 4922-5225, 4925-5204, 4935-5234, 4950-5412, 4986-5328, 5069-5353, 5080-5313, 5096-5474, 5123-5381, 5161-5646, 5177-5328, 5204-5430, 5225-5431, 5243-5514, 5251-5488, 5251-5489, 5266-5481, 5292-5732, 5294-5543, 5321-5606, 5322-5608, 5325-5786, 5415-5790, 5430-5724, 5452-5806, 5483-5715, 5504-5780, 5544-5817, 5557-5824, 5584-5824, 5587-5825, 5599-5808, 5621-5825, 5646-5823, 5663-5823
397495168CB1/ 1380	1-423, 1-481, 359-697, 367-714, 367-1024, 369-707, 369-713, 369-714, 369-1000, 376-1038, 382-779, 384-755, 389-765, 395-624, 406-714, 407-697, 431-605, 431-664, 453-876, 458-951, 469-1082, 488-900, 491-1129, 495-949, 512-1107, 512-1182, 532-1183, 536-807, 540-1125, 559-1075, 559-1183, 561-1152, 575-804, 575-1152, 582-961, 590-1163, 636-1067, 640-907, 659-926, 661-1221, 668-1252, 701-927, 702-1027, 705-1239, 724-955, 728-1017, 796-1049, 888-1327, 891-1114, 927-1121, 931-1380
407483131CB1/ 1773	1-600, 4-320, 19-510, 45-464, 53-291, 60-642, 63-336, 63-386, 163-447, 189-447, 255-567, 294-812, 294-830, 403-623, 460-737, 473-951, 475-617, 516-642, 572-750, 606-854, 616-886, 628-876, 693-987, 695-972, 737-1334, 869-1101, 870-1400, 886-1131, 925-1675, 931-1166, 931-1441, 963-1559, 992-1308, 997-1269, 1020-1296, 1038-1357, 1098-1669, 1105-1711, 1172-1422, 1176-1701, 1185-1685, 1191-1332, 1192-1709, 1192-1713, 1195-1332, 1252-1723, 1255-1723, 1258-1721, 1274-1721, 1289-1728, 1290-1500, 1290-1645, 1295-1721, 1300-1705, 1302-1773, 1306-1585, 1306-1724, 1307-1584, 1311-1447, 1311-1734, 1316-1722, 1332-1724, 1336-1724, 1338-1721, 1339-1574, 1347-1610, 1347-1721, 1347-1724, 1349-1590, 1393-1717

WO 02/097032

PCT/US02/11152

Table 4

Polynucleotide SEQ ID NO./ Incrte ID/ Sequence Length	Sequence Fragments	
41/4553860CB1/ 2810	1-444, 78-452, 79-509, 100-453, 140-799, 163-463, 359-823, 455-699, 455-722, 455-818, 455-869, 455-887, 455-931, 455-938, 455-944, 456-874, 460-1062, 464-879, 464-925, 464-985, 464-989, 464-1003, 464-1058, 466-905, 466-1129, 469-933, 473-942, 482-940, 489-880, 491-647, 491-673, 491-798, 499-877, 528-710, 642-1253, 643-1159, 842-1158, 900-1320, 907-1181, 929-1297, 930-1180, 930-1316, 955-1315, 958-1313, 983-1325, 993-1301, 998-1256, 999-1611, 1030-1290, 1064-1309, 1078-1280, 1078-1325, 1086-1283, 1107-1369, 1125-1309, 1125-1325, 1141-1325, 1332-1931, 1333-1731, 1334-1589, 1355-1624, 1361-1593, 1376-1721, 1385-1730, 1394-1797, 1394-1943, 1411-1876, 1438-1705, 1438-2022, 1439-1685, 1471-2032, 1496-1616, 1560-1834, 1600-1985, 1612-2038, 1622-2278, 1675-1928, 1779-2342, 1851-2345, 1855-2128, 1856-2119, 1915-2565, 1926-2532, 1933-2230, 1933-2261, 1942-2531, 1945-2453, 1945-2482, 1961-2531, 1963-2174, 1989-2766, 1996-2602, 1997-2488, 1999-2488, 2011-2459, 2041-2242, 2071-2300, 2086-2351, 2113-2373, 2114-2348, 2116-2368, 2116-2633, 2117-2715, 2133-2325, 2140-2363, 2146-2665, 2147-2766, 2151-2410, 2184-2731, 2259-2778, 2286-2602, 2324-2810, 2331-2465, 2344-2596, 2346-2810, 2356-2810, 2401-2474	
42/7806195CB1/ 2549	1-234, 11-454, 11-2549, 88-184, 88-310, 88-459, 88-462, 88-466, 89-519, 110-199, 110-463, 150-809, 164-462, 173-472, 279-462, 369-833, 465-709, 465-732, 465-828, 465-879, 465-897, 465-941, 465-948, 465-954, 466-884, 470-1072, 474-889, 474-935, 474-995, 474-999, 474-1013, 474-1068, 476-915, 476-1139, 479-943, 483-952, 492-950, 499-890, 501-657, 501-683, 501-808, 509-887, 538-720, 653-1159, 760-1358, 852-1157, 949-1325, 1155-1503, 1158-1366, 1158-1397, 1158-1494, 1167-1570, 1167-1716, 1184-1649, 1211-1478, 1211-1795, 1212-1458, 1244-1805, 1271-1389, 1314-1905, 1315-1722, 1315-1877, 1316-1927, 1333-1607, 1373-1758, 1381-1726, 1381-1737, 1385-1861, 1448-1701, 1508-2074, 1552-2115, 1605-1777, 1624-2118, 1628-1901, 1629-1892, 1699-2305, 1706-2003, 1706-2034, 1715-2304, 1718-2226, 1718-2255, 1734-2304, 1738-1947, 1742-2099, 1751-2253, 1762-2539, 1769-2221, 1769-2575, 1770-2261, 1772-2261, 1784-2232, 1814-2015, 1844-2073, 1859-2124, 1886-2146, 1887-2121, 1889-2141, 1889-2406, 1890-2488, 1906-2098, 1913-2136, 1920-2539, 1924-2183, 1944-2726, 1957-2504, 2022-2549, 2059-2375, 2069-2238, 2097-2549, 2104-2238, 2117-2369, 2174-2247	

WO 02/097032

PCT/US02/11152

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
22	1351608CB1	PGANNOT01
23	4259314CB1	STOMTMR02
24	3660046CB1	BRAIUNF01
25	3016416CB1	THYRDIE01
26	2133755CB1	SINTFER02
27	5259957CB1	KIDETXS02
28	55029783CB1	BLADTUT04
29	8032202CB1	TESTNOT11
30	6937367CB1	FTUBTUR01
31	3876510CB1	HEARFET02
32	4900076CB1	HNT2TXT01
33	1543848CB1	MPHGNOT03
34	6254070CB1	LUNFTUT02
35	1289839CB1	BRAHTDR04
36	5565648CB1	LIVRPET05
37	2764456CB1	COLNNOT23
38	5734806CB1	THYMNOR02
39	7495168CB1	NOSETUE01
40	7483131CB1	KIDNNOT19
41	4558650CB1	BRAUNOR01
42	7506195CB1	CONNTUT05

WO 02/097032

PCT/US02/11152

Table 6

Library	Vector	Library Description
BLADTUT04	pNCY	Library was constructed using RNA isolated from bladder tumor tissue removed from a 60-year-old Caucasian male during a radical cystectomy, prostatectomy, and vasectomy. Pathology indicated grade 3 transitional cell carcinoma in the left bladder wall. Carcinoma in-situ was identified in the dome and trigone. Patient history included tobacco use. Family history included type I diabetes, malignant neoplasm of the stomach, atherosclerotic coronary artery disease, and acute myocardial infarction.
BRAHTR04	PCDNA2.1	This random primed library was constructed using RNA isolated from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningial fibrosis and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRAUNF01	PRARE	This 5' cap isolated full-length library was constructed using RNA isolated from a DU 145 cell line derived from a brain tumor removed from a 69-year-old Caucasian male. The cells were untreated for 14 hours. Pathology indicated metastatic carcinoma. Patient history included lymphocytic leukemia for 3 years and prostate carcinoma with metastasis to the brain.
BRAUNOR01	pNCY	This random primed library was constructed using RNA isolated from striatum, globus pallidus and posterior putamen tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles.

WO 02/097032

PCT/US02/11152

Table 6

Library	Vector	Library Description
		The posterior hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive gliosis. Patient history included sepsis, cholangitis, post-operative atherosclerosis, pneumonia CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolephrosclerosis, nodular colloid goiter, emphysema, CHF, hypothyroidism, and peripheral vascular disease.
COLNNO723	pINCY	Library was constructed using RNA isolated from diseased colon tissue removed from a 16-year-old Caucasian male during a total colectomy with abdominal/perineal resection. Pathology indicated gastritis and pancolitis consistent with the acute phase of ulcerative colitis. Inflammation was more severe in the transverse colon, with inflammation confined to the mucosa. There was only mild involvement of the ascending and sigmoid colon, and no significant involvement of the cecum, rectum, or terminal ileum. Family history included irritable bowel syndrome.
CONN7UT05	pINCY	Library was constructed using RNA isolated from tumorous skull soft tissue removed from a 34-year-old Caucasian female during skull lesion excision. Pathology indicated grade 3 meningioma forming an implant in the dermis and subcutis associated with dense fibrosis. Patient history included seizures, bone cancer, and brain cancer. Surgeries included cranioplasty and cerebral meninges lesion excision, and treatment included whole brain radiation. Family history included anxiety and depression.
FTUBTUR01	PCDNA2.1	This random primed library was constructed using RNA isolated from fallopian tube tumor tissue removed from an 85-year-old Caucasian female during bilateral salpingo-oophorectomy and hysterectomy. Pathology indicated poorly differentiated mixed endometrioid (80%) and serous (20%) adenocarcinoma, which was confined to the mucosa without mural involvement. Endometrioid carcinoma <i>in situ</i> was also present. Pathology for the associated uterus tumor indicated focal endometrioid adenocarcinoma <i>in situ</i> and moderately differentiated invasive adenocarcinoma arising in an endometrial polyp. Metastatic endometrioid and serous adenocarcinoma was present at the cut-de-sac tumor. Patient history included mediastinal carcinoma of the thyroid and myocardial infarction.
HEARFET02	pINCY	Library was constructed using RNA isolated from heart tissue removed from a Caucasian male fetus, who was stillborn with a hypoplastic left heart and died at 23 weeks' gestation.
HNT2TX01	pINCY	Library was constructed using RNA isolated from the hNT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor). Cells were treated with mouse leptin at 1 ug/ml and 9 cis retinoic acid at 3.3 uM for 6 days.

WO 02/097032

PCT/US02/11152

Table 6

Library	Vector	Library Description
KIDBTXS02	pNCY	This subcloned, transformed embryonal cell line library was constructed using 9 million clones from a treated, transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue and was subjected to two rounds of subtraction hybridization with 1.9 million clones from an untreated transformed embryonal cell line (293-EBNA) derived from a kidney epithelial tissue library. The starting library for subtraction was constructed using RNA isolated from the treated, transformed embryonal cell line (293-EBNA). The cells were treated with 5-aza-2'-deoxycytidine and transformed with adenovirus 5 DNA. The hybridization probe for subtraction was derived from a similarly constructed library from RNA isolated from untreated 293-EBNA cells from the same cell line. Subtractive hybridization conditions were based on the methodologies of Swarcop et al., NAR 19 (1991):1954 and Boudado, et al. Genome Research (1996) 6:791.
KIDNNOT19	pNCY	Library was constructed using RNA isolated from kidney tissue removed a 65-year-old Caucasian male during an exploratory laparotomy and nephroureterectomy. Pathology for the associated tumor tissue indicated a grade 1 renal cell carcinoma within the upper pole of the left kidney. Patient history included malignant melanoma of the abdominal skin, benign neoplasm of colon, cerebrovascular disease, and umbilical hernia. Family history included myocardial infarction, atherosclerotic coronary artery disease, cerebrovascular disease, prostate cancer, myocardial infarction, and atherosclerotic coronary artery disease.
LIVRHE05	pNCY	Library was constructed using RNA isolated from CD34+ progenitor cell tissue removed from the liver of a Caucasian male fetus who died after 20 weeks' gestation.
LUNPTUT02	pNCY	Library was constructed using RNA isolated from pleura tumor tissue removed from a 55-year-old Caucasian female during complete pneumonectomy. Pathology indicated grade 3 sarcoma most consistent with leiomyosarcoma, uterine primary, forming a bosselated mass replacing the right lower lobe and a portion of the middle lobe. The tumor involved the adjacent parietal pleura and pericardium. Multiple nodules comprising the tumor show near total necrosis. The right upper lobe was atelectic but uninvolved by tumor. Microsections of cellular nodules show brisk mitotic activity. The pericardium shows direct involvement but its margins were tumor free. Smooth muscle actin was positive. Estrogen receptor was negative and progesterone receptor was positive. Patient history included shortness of breath, peptic ulcer disease, lung cancer, uterine cancer, normal delivery, tobacco abuse, and deficiency anemia. Previous surgeries included endoscopic excision of a lung lesion. Family history included atherosclerotic coronary artery disease, breast cancer, type II diabetes, and multiple sclerosis.
MPHONOT03	PBLUESCRIPT	Library was constructed using RNA isolated from plastic adherent mononuclear cells isolated from buffy coat units obtained from unrelated male and female donors.

WO 02/097032

PCT/US02/11152

Table 6

Library	Vector	Library Description
NOSUTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from nasal and cribriform tumor tissue removed from a 65-year-old Caucasian male during total face osteotomy with reconstruction, rhinotomy and craniotomy. Pathology indicated olfactory neuroblastoma in the nasal cavity and cribriform region. The patient presented with cancer of the head, face and neck, and epistaxis. Patient history included extrinsic asthma, cancer of the head, face and neck, and epistaxis. Previous surgeries included total face osteotomy with reconstruction. Patient medications included Biaxin, Aleve, and Valium. The patient received radiation treatments. Family history included chronic sinusitis in the mother and type II diabetes in the father.
PGANNOT01	PSPORT1	Library was constructed using RNA isolated from paraganglionic tumor tissue removed from the intra-abdominal region of a 46-year-old Caucasian male during exploratory laparotomy. Pathology indicated a benign paraganglioma and was associated with a grade 2 renal cell carcinoma, clear cell type, which did not penetrate the capsule. Surgical margins were negative for tumor.
SINTEP02	pINCY	This random primed library was constructed using RNA isolated from small intestine tissue removed from a Caucasian male fetus who died from fetal demise.
STOMTMR02	PCDNA2.1	This random primed library was constructed using RNA isolated from diseased stomach tissue removed from a 76-year-old Caucasian male during proximal gastrectomy and partial esophagectomy. Pathology indicated chronic gastritis. Pathology for the matched tumor tissue indicated invasive grade 3 adenocarcinoma forming an ulcerated mass at the gastro-esophageal junction. The tumor infiltrated through the muscularis propria into the periesophageal adipose tissue. One of four perigastric lymph nodes was positive for tumor. Patient history included dysphagia, atherosclerotic coronary artery disease, malignant melanoma of the skin, COPD, benign neoplasm of the large bowel, malignant neoplasm of upper lobe of lung, and alcohol abuse. Family history included atherosclerotic coronary artery disease and myocardial infarction.
TESTNOT11	pINCY	Library was constructed using RNA isolated from testicular tissue removed from a 16-year-old Caucasian male who died from hanging. Patient history included drug use (tobacco, marijuana, and cocaine use), and medications included Lithium, Ritalin, and Paxil.
THYMNOR02	pINCY	The library was constructed using RNA isolated from thymus tissue removed from a 2-year-old Caucasian female during a thymectomy and patch closure of left arioventricular fistula. Pathology indicated there was no gross abnormality of the thymus. The patient presented with congenital heart abnormalities. Patient history included double inlet left ventricle and a rudimentary right ventricle, pulmonary hypertension, cyanosis, subaortic stenosis, seizures, and a fracture of the skull base. Family history included reflux neuropathy.

WO 02/097032

PCT/US02/11152

Table 6

Library	Vector	Library Description
THYRDIE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from diseased thyroid tissue removed from a 22-year-old Caucasian female during closed thyroid biopsy, partial thyroidectomy, and regional lymph node excision. Pathology indicated adenomatous hyperplasia. The patient presented with malignant neoplasm of the thyroid. Patient history included normal delivery, alcohol abuse, and tobacco abuse. Previous surgeries included myringotomy. Patient medications included an unspecified type of birth control pills. Family history included hyperlipidemia and depressive disorder in the mother; and benign hypertension, congestive heart failure, and chronic leukemia in the grandparent(s).

WO 02/097032

PCT/US02/11152

Table 7

Program	Description	Reference	Parameter Threshold
ABI/FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, blastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less; Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, fasta, tfastx, and search.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:462-489.	ESTs: fasta E value=1.0E-6; Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less; Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less

WO 02/097032

PCT/US02/11152

Table 7

Program	Description	Reference	Parameter Threshold
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, SMART and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:501-531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM, SMART or TIGRFAM hits: Probability value= 1.0E-3 or less: Signal peptide hits: Score= 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribakov, M. et al. (1988) CABIOS 4:61-66; Gribakov, M. et al. (1989) Methods Enzymol. 183:146-159; Baroch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score=GCC-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nelson, H. et al. (1997) Protein Engineering 10:1-6; Clavette, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:563-571.	

WO 02/097032

PCT/US02/11152

Table 7

Program	Description	Reference	Parameter Threshold
TMHMM	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sommadossi, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

WO 02/097032

PCT/US02/11152

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
42	7506195	675516731	SNP00121036	111	584	A	A	G	Q145	n/d	n/a	n/a	n/a
42	7506195	676302411	SNP00121036	111	584	A	A	G	Q145	n/d	n/a	n/a	n/a
42	7506195	676744911	SNP00121036	110	584	G	A	G	R145	n/d	n/a	n/a	n/a
42	7506195	676902911	SNP00121036	109	584	G	A	G	R145	n/d	n/a	n/a	n/a
42	7506195	775432111	SNP00121036	215	420	A	A	G	P90	n/d	n/a	n/a	n/a

WO 02/097032

PCT/US02/11152

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
 - a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21,
 - 5 b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:3-4, SEQ ID NO:6, SEQ ID NO:8-9, SEQ ID NO:11-12, SEQ ID NO:14-16, SEQ ID NO:18-21,
 - 10 c) a polypeptide comprising a naturally occurring amino acid sequence at least 93% identical to the amino acid sequence of SEQ ID NO:1,
 - d) a polypeptide comprising a naturally occurring amino acid sequence at least 95% identical to the amino acid sequence of SEQ ID NO:2,
 - e) a polypeptide comprising a naturally occurring amino acid sequence at least 96% identical to the amino acid sequence of SEQ ID NO:5,
 - 15 f) a polypeptide comprising a naturally occurring amino acid sequence at least 98% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:7 and SEQ ID NO:10,
 - g) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to the polynucleotide sequence of SEQ ID NO:17,
 - 20 h) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and
 - i) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.
- 25 2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.
3. An isolated polynucleotide encoding a polypeptide of claim 1.
- 30 4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42.

WO 02/097032

PCT/US02/11152

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

- a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
- b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42,
- b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:32-37, SEQ ID NO:39-42,
- c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 92% identical to the polynucleotide sequence of SEQ ID NO:26,
- d) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 98% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28 and SEQ ID NO:31,
- e) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 99% identical to the polynucleotide sequence of SEQ ID NO:38,

WO 02/097032

PCT/US02/11152

- f) a polynucleotide complementary to a polynucleotide of a),
 g) a polynucleotide complementary to a polynucleotide of b),
 h) a polynucleotide complementary to a polynucleotide of c),
 i) a polynucleotide complementary to a polynucleotide of d),
 5 j) a polynucleotide complementary to a polynucleotide of e), and
 k) an RNA equivalent of a)-j).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

10

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and
 15 which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
 b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

20

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- 25 a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
 b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

30 17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence

WO 02/097032

PCT/US02/11152

selected from the group consisting of SEQ ID NO:1-21.

19. A method for treating a disease or condition associated with decreased expression of functional CGDD, comprising administering to a patient in need of such treatment the composition of
5 claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- 10 a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

15 22. A method for treating a disease or condition associated with decreased expression of functional CGDD, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- 20 a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and
25 a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional CGDD, comprising administering to a patient in need of such treatment a composition of claim 24.

30 26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

WO 02/097032

PCT/US02/11152

- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:
- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
 - b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
 - c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:
- a) treating a biological sample containing nucleic acids with the test compound,
 - b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,
 - c) quantifying the amount of hybridization complex, and
 - d) comparing the amount of hybridization complex in the treated biological sample with

WO 02/097032

PCT/US02/11152

the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

- 5 30. A diagnostic test for a condition or disease associated with the expression of CGDD in a biological sample, the method comprising:
- a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and
 - 10 b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.
31. The antibody of claim 11, wherein the antibody is:
- a) a chimeric antibody,
 - 15 b) a single chain antibody,
 - c) a Fab fragment,
 - d) a F(ab')₂ fragment, or
 - e) a humanized antibody.
- 20 32. A composition comprising an antibody of claim 11 and an acceptable excipient.
33. A method of diagnosing a condition or disease associated with the expression of CGDD in a subject, comprising administering to said subject an effective amount of the composition of claim 32.
- 25 34. A composition of claim 32, wherein the antibody is labeled.
35. A method of diagnosing a condition or disease associated with the expression of CGDD in a subject, comprising administering to said subject an effective amount of the composition of claim 34.
- 30 36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

WO 02/097032

PCT/US02/11152

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibodies from said animal, and
- 5 c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.
37. A polyclonal antibody produced by a method of claim 36.
- 10 38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.
39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:
- 15 a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibody producing cells from the animal,
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- 20 d) culturing the hybridoma cells, and
- e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.
- 25 40. A monoclonal antibody produced by a method of claim 39.
41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.
- 30 42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.
43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant

WO 02/097032

PCT/US02/11152

immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21 in a sample, the method comprising:

- 5 a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21 in the sample.

10

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21 from a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- 15 b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

- a) labeling the polynucleotides of the sample,
- 25 b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and
- c) quantifying the expression of the polynucleotides in the sample.

30 48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

WO 02/097032

PCT/US02/11152

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is
5 completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

10 52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

15 54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical
20 location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

25 56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

30 59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

WO 02/097032

PCT/US02/11152

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.
62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.
- 5 63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.
64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.
65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.
- 10 66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.
67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.
- 15 68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.
69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.
70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.
- 20 71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.
72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.
- 25 73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.
74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.
75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.
- 30 76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.
77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.

WO 02/097032

PCT/US02/11152

78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

5 80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:25.

81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.

82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.

10 83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.

84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.

15 85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.

86. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:31.

87. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:32.

20 88. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:33.

89. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:34.

25 90. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:35.

91. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:36.

92. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:37.

30 93. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:38.

94. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:39.

WO 02/097032

PCT/US02/11152

95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:40.

96. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:41.

5 97. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:42.

WO 02/097032

PCT/US02/11152

<110> INCYTE GENOMICS, INC.

AZIMZAI, Yalda
AU-YOUNG, Janice K.
BATRA, Sajeev
BAUGHN, Mariah R.
BECHA, Shanya D.
BOROWSKY, Mark L.
BUPFORD, Neil
DING, Li
ELLIOTT, Vicki S.
EMERLING, Brooke M.
GANDHI, Ameena R.
GLEITZEN, Kimberly J.
GRIFPIN, Jennifer A.
HAFALIA, April J.A.
HONCHELL, Cynthia D.
LAL, Preeti G.
LEE, Soo Yeun
LIU, Dyung Aina M.
ARVIZU, Chandra S.
RAMKUMAR, Jayalaxmi
REDDY, Roopa
SANJANWALA, Madhu, M.
TANG, Y. Tom
WALLA, Narinder K.
WANG, Yu-mei, E.
WARREN, Bridget A.
XU, Yuming
YANG, Junming
YAO, Monique G.
YUE, Henry
ZEBARJADIAN, Yeganeh

<120> PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH

<130> PI-0417 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/282,110; 60/283,294; 60/286,820; 60/287,228;
60/291,662; 60/291,846; 60/293,727; 60/295,340;
60/295,263; 60/349,705

<151> 2001-04-06; 2001-04-11; 2001-04-26; 2001-04-27;
2001-05-16; 2001-05-18; 2001-05-25; 2001-06-01;
2001-06-01; 2002-01-15

<160> 42

<170> PERL Program

<210> 1

<211> 1738

<212> FRT

<213> Homo sapiens

WO 02/097032

PCT/US02/11152

<220>

<221> misc_feature

<223> Incyte ID No: 1351608CD1

<400> 1

```

Met Glu Ile Ser Ala Glu Leu Pro Gln Thr Pro Gln Arg Leu Ala
 1      5      10      15
Ser Trp Trp Asp Gln Gln Val Asp Phe Tyr Thr Ala Phe Leu His
 20      25      30
His Leu Ala Gln Leu Val Pro Glu Ile Tyr Phe Ala Glu Met Asp
 35      40      45
Pro Asp Leu Glu Lys Gln Glu Glu Ser Val Gln Met Ser Ile Phe
 50      55      60
Thr Pro Leu Glu Trp Tyr Leu Phe Gly Glu Asp Pro Asp Ile Cys
 65      70      75
Leu Glu Lys Leu Lys His Ser Gly Ala Phe Gln Leu Cys Gly Arg
 80      85      90
Val Phe Lys Ser Gly Glu Thr Thr Tyr Ser Cys Arg Asp Cys Ala
 95     100     105
Ile Asp Pro Thr Cys Val Leu Cys Met Asp Cys Phe Gln Asp Ser
110     115     120
Val His Lys Asn His Arg Tyr Lys Met His Thr Ser Thr Gly Gly
125     130     135
Gly Phe Cys Asp Cys Gly Asp Thr Glu Ala Trp Lys Thr Gly Pro
140     145     150
Phe Cys Val Asn His Glu Pro Gly Arg Ala Gly Thr Ile Lys Glu
155     160     165
Asn Ser Arg Cys Pro Leu Asn Glu Glu Val Ile Val Gln Ala Arg
170     175     180
Lys Ile Phe Pro Ser Val Ile Lys Tyr Val Val Glu Met Thr Ile
185     190     195
Trp Glu Glu Glu Lys Glu Leu Pro Pro Glu Leu Gln Ile Arg Glu
200     205     210
Lys Asn Glu Arg Tyr Tyr Cys Val Leu Phe Asn Asp Glu His His
215     220     225
Ser Tyr Asp His Val Ile Tyr Ser Leu Gln Arg Ala Leu Asp Cys
230     235     240
Glu Leu Ala Glu Ala Gln Leu His Thr Thr Ala Ile Asp Lys Glu
245     250     255
Gly Arg Arg Ala Val Lys Ala Gly Ala Tyr Ala Ala Cys Gln Glu
260     265     270
Ala Lys Glu Asp Ile Lys Ser His Ser Glu Asn Val Ser Gln His
275     280     285
Pro Leu His Val Glu Val Leu His Ser Glu Ile Met Ala His Gln
290     295     300
Lys Phe Ala Leu Arg Leu Gly Ser Trp Met Asn Lys Ile Met Ser
305     310     315
Tyr Ser Ser Asp Phe Arg Gln Ile Phe Cys Gln Ala Cys Leu Arg
320     325     330
Glu Glu Pro Asp Ser Glu Asn Pro Cys Leu Ile Ser Arg Leu Met
335     340     345
Leu Trp Asp Ala Lys Leu Tyr Lys Gly Ala Arg Lys Ile Leu His
350     355     360
Glu Leu Ile Phe Ser Ser Phe Phe Met Glu Met Glu Tyr Lys Lys
365     370     375
Leu Phe Ala Met Glu Phe Val Lys Tyr Tyr Lys Gln Leu Gln Lys

```

WO 02/097032

PCT/US02/11152

```

380          385          390
Glu Tyr Ile Ser Asp Asp His Asp Arg Ser Ile Ser Ile Thr Ala
395          400          405
Leu Ser Val Gln Met Phe Thr Val Pro Thr Leu Ala Arg His Leu
410          415          420
Ile Glu Glu Gln Asn Val Ile Ser Val Ile Thr Glu Thr Leu Leu
425          430          435
Glu Val Leu Pro Glu Tyr Leu Asp Arg Asn Asn Lys Phe Asn Phe
440          445          450
Gln Gly Tyr Ser Gln Asp Lys Leu Gly Arg Val Tyr Ala Val Ile
455          460          465
Cys Asp Leu Lys Tyr Ile Leu Ile Ser Lys Pro Thr Ile Trp Thr
470          475          480
Glu Arg Leu Arg Met Gln Phe Leu Glu Gly Phe Arg Ser Phe Leu
485          490          495
Lys Ile Leu Thr Cys Met Gln Gly Met Glu Glu Ile Arg Arg Gln
500          505          510
Val Gly Gln His Ile Glu Val Asp Pro Asp Trp Glu Ala Ala Ile
515          520          525
Ala Ile Gln Met Gln Leu Lys Asn Ile Leu Leu Met Phe Gln Glu
530          535          540
Trp Cys Ala Cys Asp Glu Glu Leu Leu Leu Val Ala Tyr Lys Glu
545          550          555
Cys His Lys Ala Val Met Arg Cys Ser Thr Ser Phe Ile Ser Ser
560          565          570
Ser Lys Thr Val Val Gln Ser Cys Gly His Ser Leu Glu Thr Lys
575          580          585
Ser Tyr Arg Val Ser Glu Asp Leu Val Ser Ile His Leu Pro Leu
590          595          600
Ser Arg Thr Leu Ala Gly Leu His Val Arg Leu Ser Arg Leu Gly
605          610          615
Ala Val Ser Arg Leu His Glu Phe Val Ser Phe Glu Asp Phe Gln
620          625          630
Val Glu Val Leu Val Glu Tyr Pro Leu Arg Cys Leu Val Leu Val
635          640          645
Ala Gln Val Val Ala Glu Met Trp Arg Arg Asn Gly Leu Ser Leu
650          655          660
Ile Ser Gln Val Phe Tyr Tyr Gln Asp Val Lys Cys Arg Glu Glu
665          670          675
Met Tyr Asp Lys Asp Ile Ile Met Leu Gln Ile Gly Ala Ser Leu
680          685          690
Met Asp Pro Asn Lys Phe Leu Leu Leu Val Leu Gln Arg Tyr Glu
695          700          705
Leu Ala Glu Ala Phe Asn Lys Thr Ile Ser Thr Lys Asp Gln Asp
710          715          720
Leu Ile Lys Gln Tyr Asn Thr Leu Ile Glu Glu Met Leu Gln Val
725          730          735
Leu Ile Tyr Ile Val Gly Glu Arg Tyr Val Pro Gly Val Gly Asn
740          745          750
Val Thr Lys Glu Glu Val Thr Met Arg Glu Ile Ile His Leu Leu
755          760          765
Cys Ile Glu Pro Met Pro His Ser Ala Ile Ala Lys Asn Leu Pro
770          775          780
Glu Asn Glu Asn Asn Glu Thr Gly Leu Glu Asn Val Ile Asn Lys
785          790          795
Val Ala Thr Phe Lys Lys Pro Gly Val Ser Gly His Gly Val Tyr

```


WO 02/097032

PCT/US02/11152

	800	805	810
Glu Leu Lys Asp	Glu Ser Leu Lys Asp	Phe Asn Met Tyr Phe Tyr	
815	820	825	
His Tyr Ser Lys	Thr Gln His Ser Lys	Ala Glu His Met Gln Lys	
830	835	840	
Lys Arg Arg Lys	Gln Glu Asn Lys Asp	Glu Ala Leu Pro Pro Pro	
845	850	855	
Pro Pro Pro Glu	Phe Cys Pro Ala Phe	Ser Lys Val Ile Asn Leu	
860	865	870	
Leu Asn Cys Asp	Ile Met Met Tyr Ile	Leu Arg Thr Val Phe Glu	
875	880	885	
Arg Ala Ile Asp	Thr Asp Ser Asn Leu Trp	Thr Glu Gly Met Leu	
890	895	900	
Gln Met Ala Phe	His Ile Leu Ala Leu	Gly Leu Leu Glu Glu Lys	
905	910	915	
Gln Gln Leu Gln	Lys Ala Pro Glu Glu	Glu Val Thr Phe Asp Phe	
920	925	930	
Tyr His Lys Ala	Ser Arg Leu Gly Ser	Ser Ala Met Asn Ile Gln	
935	940	945	
Met Leu Leu Glu	Lys Leu Lys Gly Ile	Pro Gln Leu Glu Gly Gln	
950	955	960	
Lys Asp Met Ile	Thr Trp Ile Leu Gln	Met Phe Asp Thr Val Lys	
965	970	975	
Arg Leu Arg Glu	Lys Ser Cys Leu Ile	Val Ala Thr Thr Ser Gly	
980	985	990	
Ser Glu Ser Ile	Lys Asn Asp Glu Ile	Thr His Asp Lys Glu Lys	
995	1000	1005	
Ala Glu Arg Lys	Arg Lys Ala Glu Ala	Ala Arg Leu His Arg Gln	
1010	1015	1020	
Lys Ile Met Ala	Gln Met Ser Ala Leu	Gln Lys Asn Phe Ile Glu	
1025	1030	1035	
Thr His Lys Leu	Met Tyr Asp Asn Thr	Ser Glu Met Pro Gly Lys	
1040	1045	1050	
Glu Asp Ser Ile	Met Glu Glu Glu Ser	Thr Pro Ala Val Ser Asp	
1055	1060	1065	
Tyr Ser Arg Ile	Ala Leu Gly Pro Lys	Arg Gly Pro Ser Val Thr	
1070	1075	1080	
Glu Lys Glu Val	Leu Thr Cys Ile Leu	Cys Gln Glu Glu Gln Glu	
1085	1090	1095	
Val Lys Ile Glu	Asn Asn Ala Met Val	Leu Ser Ala Cys Val Gln	
1100	1105	1110	
Lys Ser Thr Ala	Leu Thr Gln His Arg	Gly Lys Pro Ile Glu Leu	
1115	1120	1125	
Ser Gly Glu Ala	Leu Asp Pro Leu Phe	Met Asp Pro Asp Leu Ala	
1130	1135	1140	
Tyr Gly Thr Tyr	Thr Gly Ser Cys Gly	His Val Met His Ala Val	
1145	1150	1155	
Cys Trp Gln Lys	Tyr Phe Glu Ala Val	Gln Leu Ser Ser Gln Gln	
1160	1165	1170	
Arg Ile His Val	Asp Leu Phe Asp Leu	Glu Ser Gly Glu Tyr Leu	
1175	1180	1185	
Cys Pro Leu Cys	Lys Ser Leu Cys Asn	Thr Val Ile Pro Ile Ile	
1190	1195	1200	
Pro Leu Gln Pro	Gln Lys Ile Asn Ser	Glu Asn Ala Asp Ala Leu	
1205	1210	1215	
Ala Gln Leu Leu	Thr Leu Ala Arg Trp	Ile Gln Thr Val Leu Ala	

WO 02/097032

PCT/US02/11152

```

1220      1225      1230
Arg Ile Ser Gly Tyr Asn Ile Arg His Ala Lys Gly Glu Asn Pro
1235      1240      1245
Ile Pro Ile Phe Phe Asn Gln Gly Met Gly Asp Ser Thr Leu Glu
1250      1255      1260
Phe His Ser Ile Leu Ser Phe Gly Val Glu Ser Ser Ile Lys Tyr
1265      1270      1275
Ser Asn Ser Ile Lys Glu Met Val Ile Leu Phe Ala Thr Thr Ile
1280      1285      1290
Tyr Arg Ile Gly Leu Lys Val Pro Pro Asp Glu Arg Asp Pro Arg
1295      1300      1305
Val Pro Met Leu Thr Trp Ser Thr Cys Ala Phe Thr Ile Gln Ala
1310      1315      1320
Ile Glu Asn Leu Leu Gly Asp Glu Gly Lys Pro Leu Phe Gly Ala
1325      1330      1335
Leu Gln Asn Arg Gln His Asn Gly Leu Lys Ala Leu Met Gln Phe
1340      1345      1350
Ala Val Ala Gln Arg Ile Thr Cys Pro Gln Val Leu Ile Gln Lys
1355      1360      1365
His Leu Val Arg Leu Leu Ser Val Val Leu Pro Asn Ile Lys Ser
1370      1375      1380
Glu Asp Thr Pro Cys Leu Leu Ser Ile Asp Leu Phe His Val Leu
1385      1390      1395
Val Gly Ala Val Leu Ala Phe Pro Ser Leu Tyr Trp Asp Asp Pro
1400      1405      1410
Val Asp Leu Gln Pro Ser Ser Val Ser Ser Ser Tyr Asn His Leu
1415      1420      1425
Tyr Leu Phe His Leu Ile Thr Met Ala His Met Leu Gln Ile Leu
1430      1435      1440
Leu Thr Val Asp Thr Gly Leu Pro Leu Ala Gln Val Gln Glu Asp
1445      1450      1455
Ser Glu Glu Ala His Ser Ala Ser Ser Phe Phe Ala Glu Ile Ser
1460      1465      1470
Gln Tyr Thr Ser Gly Ser Ile Gly Cys Asp Ile Pro Gly Trp Tyr
1475      1480      1485
Leu Trp Val Ser Leu Lys Asn Gly Ile Thr Pro Tyr Leu Arg Cys
1490      1495      1500
Ala Ala Leu Phe Phe His Tyr Leu Leu Gly Val Thr Pro Pro Glu
1505      1510      1515
Glu Leu His Thr Asn Ser Ala Glu Gly Glu Tyr Ser Ala Leu Cys
1520      1525      1530
Ser Tyr Leu Ser Leu Pro Thr Asn Leu Phe Leu Leu Phe Gln Glu
1535      1540      1545
Tyr Trp Asp Thr Val Arg Pro Leu Leu Gln Arg Trp Cys Ala Asp
1550      1555      1560
Pro Ala Leu Leu Asn Cys Leu Lys Gln Lys Asn Thr Val Val Arg
1565      1570      1575
Tyr Pro Arg Lys Arg Asn Ser Leu Ile Glu Leu Pro Asp Asp Tyr
1580      1585      1590
Ser Cys Leu Leu Asn Gln Ala Ser His Phe Arg Cys Pro Arg Ser
1595      1600      1605
Ala Asp Asp Glu Arg Lys His Pro Val Leu Cys Leu Phe Cys Gly
1610      1615      1620
Ala Ile Leu Cys Ser Gln Asn Ile Cys Cys Gln Glu Ile Val Asn
1625      1630      1635
Gly Glu Glu Val Gly Ala Cys Ile Phe His Ala Leu His Cys Gly

```

WO 02/097032

PCT/US02/11152

1640 1645 1650
Ala Gly Val Cys Ile Phe Leu Lys Ile Arg Glu Cys Arg Val Val
1655 1660 1665
Leu Val Glu Gly Lys Ala Arg Gly Cys Ala Tyr Pro Ala Pro Tyr
1670 1675 1680
Leu Asp Glu Tyr Gly Glu Thr Asp Pro Gly Leu Lys Arg Gly Asn
1685 1690 1695
Pro Leu His Leu Ser Arg Glu Arg Tyr Arg Lys Leu His Leu Val
1700 1705 1710
Trp Gln Gln His Cys Ile Ile Glu Glu Ile Ala Arg Ser Gln Glu
1715 1720 1725
Thr Asn Gln Met Leu Phe Gly Phe Asn Trp Gln Leu Leu
1730 1735

<210> 2
<211> 389
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4259314CD1

<400> 2
Met Glu Gly Ser Glu Pro Val Ala Ala His Gln Gly Glu Glu Ala
1 5 10 15
Ser Cys Ser Ser Trp Gly Thr Gly Ser Thr Asn Lys Asn Leu Pro
20 25 30
Ile Met Ser Thr Ala Ser Val Glu Ile Asp Asp Ala Leu Tyr Ser
35 40 45
Arg Gln Arg Tyr Val Leu Gly Asp Thr Ala Met Gln Lys Met Ala
50 55 60
Lys Ser His Val Phe Leu Ser Gly Met Gly Gly Leu Gly Leu Glu
65 70 75
Ile Ala Lys Asn Leu Val Leu Ala Gly Ile Lys Ala Val Thr Ile
80 85 90
His Asp Thr Glu Lys Cys Gln Ala Trp Asp Leu Gly Thr Asn Phe
95 100 105
Phe Leu Ser Glu Asp Asp Val Val Asn Lys Arg Asn Arg Ala Glu
110 115 120
Ala Val Leu Lys His Ile Ala Glu Leu Asn Pro Tyr Val His Val
125 130 135
Thr Ser Ser Ser Val Pro Phe Asn Glu Thr Thr Asp Leu Ser Phe
140 145 150
Leu Asp Lys Tyr Gln Cys Val Val Leu Thr Glu Met Lys Leu Pro
155 160 165
Leu Gln Lys Lys Ile Asn Asp Phe Cys Arg Ser Gln Cys Pro Pro
170 175 180
Ile Lys Phe Ile Ser Ala Asp Val His Gly Ile Trp Ser Arg Leu
185 190 195
Phe Cys Asp Phe Gly Asp Glu Phe Glu Val Leu Asp Thr Thr Gly
200 205 210
Glu Glu Pro Lys Glu Ile Phe Ile Ser Asn Ile Thr Gln Ala Asn
215 220 225
Pro Gly Ile Val Thr Cys Leu Glu Asn His Pro His Lys Leu Glu
230 235 240

WO 02/097032

PCT/US02/11152

```

Thr Gly Gln Phe Leu Thr Phe Arg Glu Ile Asn Gly Met Thr Gly
245 250 255
Leu Asn Gly Ser Ile Gln Gln Ile Thr Val Ile Ser Pro Phe Ser
260 265 270
Phe Ser Ile Gly Asp Thr Thr Glu Leu Glu Pro Tyr Leu His Gly
275 280 285
Gly Ile Ala Val Gln Val Lys Thr Pro Lys Thr Val Phe Phe Glu
290 295 300
Ser Leu Glu Arg Gln Leu Lys His Pro Lys Cys Leu Ile Val Asp
305 310 315
Phe Ser Asn Pro Glu Ala Pro Leu Glu Ile His Thr Ala Met Leu
320 325 330
Ala Leu Asp Gln Phe Gln Glu Lys Tyr Ser Arg Lys Pro Asn Val
335 340 345
Gly Cys Gln Gln Asp Ser Glu Glu Leu Leu Lys Leu Ala Thr Ser
350 355 360
Ile Ser Glu Thr Leu Glu Glu Lys Val Thr Ile Glu Ile Tyr Gly
365 370 375
Cys Pro Asn Ile Cys Leu Leu Ile His Lys Cys Ser Val Tyr
380 385

```

<210> 3

<211> 854

<212> FRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3660046CD1

<400> 3

```

Met Glu Arg Pro Tyr Thr Phe Lys Asp Phe Leu Leu Arg Pro Arg
1 5 10 15
Ser His Lys Ser Arg Val Lys Gly Phe Leu Arg Leu Lys Met Ala
20 25 30
Tyr Met Pro Lys Asn Gly Gly Gln Asp Glu Glu Asn Ser Asp Gln
35 40 45
Arg Asp Asp Met Glu His Gly Trp Glu Val Val Asp Ser Asn Asp
50 55 60
Ser Ala Ser Gln His Gln Glu Glu Leu Pro Pro Pro Pro Leu Pro
65 70 75
Pro Gly Trp Glu Glu Lys Val Asp Asn Leu Gly Arg Thr Tyr Tyr
80 85 90
Val Asn His Asn Asn Arg Thr Thr Gln Trp His Arg Pro Ser Leu
95 100 105
Met Asp Val Ser Ser Glu Ser Asp Asn Asn Ile Arg Gln Ile Asn
110 115 120
Gln Glu Ala Ala His Arg Arg Phe Arg Ser Arg Arg His Ile Ser
125 130 135
Glu Asp Leu Glu Pro Glu Pro Ser Glu Gly Gly Asp Val Pro Glu
140 145 150
Pro Trp Glu Thr Ile Ser Glu Glu Val Asn Ile Ala Gly Asp Ser
155 160 165
Leu Gly Leu Ala Leu Pro Pro Pro Pro Ala Ser Pro Gly Ser Arg
170 175 180
Thr Ser Pro Gln Glu Leu Ser Glu Glu Leu Ser Arg Arg Leu Gln

```

WO 02/097032

PCT/US02/11152

	185		190		195
Ile Thr Pro Asp	Ser Asn Gly Glu Gln Phe Ser Ser Leu Ile Gln				
	200		205		210
Arg Glu Pro Ser	Ser Arg Leu Arg Ser Cys Ser Val Thr Asp Ala				
	215		220		225
Val Ala Glu Gln Gly	His Leu Pro Pro Ser Ala Pro Ala Gly				
	230		235		240
Arg Ala Arg Ser	Ser Thr Val Thr Gly Gly Glu Glu Pro Thr Pro				
	245		250		255
Ser Val Ala Tyr	Val His Thr Thr Pro Gly Leu Pro Ser Gly Trp				
	260		265		270
Glu Glu Arg Lys	Asp Ala Lys Gly Arg Thr Tyr Tyr Val Asn His				
	275		280		285
Asn Asn Arg Thr	Thr Thr Trp Thr Arg Pro Ile Met Gln Leu Ala				
	290		295		300
Glu Asp Gly Ala	Ser Gly Ser Ala Thr Asn Ser Asn Asn His Leu				
	305		310		315
Ile Glu Pro Gln	Ile Arg Arg Pro Arg Ser Leu Ser Ser Pro Thr				
	320		325		330
Val Thr Leu Ser	Ala Pro Leu Glu Gly Ala Lys Asp Ser Pro Val				
	335		340		345
Arg Arg Ala Val	Lys Asp Thr Leu Ser Asn Pro Gln Ser Pro Gln				
	350		355		360
Pro Ser Pro Tyr	Asn Ser Pro Lys Pro Gln His Lys Val Thr Gln				
	365		370		375
Ser Phe Leu Pro	Pro Gly Trp Glu Met Arg Ile Ala Pro Asn Gly				
	380		385		390
Arg Pro Phe Phe	Ile Asp His Asn Thr Lys Thr Thr Thr Trp Glu				
	395		400		405
Asp Pro Arg Leu	Lys Phe Pro Val His Met Arg Ser Lys Thr Ser				
	410		415		420
Leu Asn Pro Asn	Asp Leu Gly Pro Leu Pro Pro Gly Trp Glu Glu				
	425		430		435
Arg Thr His Thr	Asp Gly Arg Ile Phe Tyr Ile Asn His Asn Ile				
	440		445		450
Lys Arg Thr Gln	Trp Glu Asp Pro Arg Leu Glu Asn Val Ala Ile				
	455		460		465
Thr Gly Pro Ala	Val Pro Tyr Ser Arg Asp Tyr Lys Arg Lys Tyr				
	470		475		480
Glu Phe Phe Arg	Arg Lys Leu Lys Lys Gln Asn Asp Ile Pro Asn				
	485		490		495
Lys Phe Glu Met	Lys Leu Arg Arg Ala Thr Val Leu Glu Asp Ser				
	500		505		510
Tyr Arg Arg Ile	Met Gly Val Lys Arg Ala Asp Phe Leu Lys Ala				
	515		520		525
Arg Leu Trp Ile	Glu Phe Asp Gly Glu Lys Gly Leu Asp Tyr Gly				
	530		535		540
Gly Val Ala Arg	Glu Trp Phe Phe Leu Ile Ser Lys Glu Met Phe				
	545		550		555
Asn Pro Tyr Tyr	Gly Leu Phe Glu Tyr Ser Ala Thr Asp Asn Tyr				
	560		565		570
Thr Leu Gln Ile	Asn Pro Asn Ser Gly Leu Cys Asn Glu Asp His				
	575		580		585
Leu Ser Tyr Phe	Lys Phe Ile Gly Arg Val Ala Gly Met Ala Val				
	590		595		600
Tyr His Gly Lys	Leu Leu Asp Gly Phe Phe Ile Arg Pro Phe Tyr				

WO 02/097032

PCT/US02/11152

```

605          610          615
Lys Met Met Leu His Lys Pro Ile Thr Leu His Asp Met Glu Ser
620          625          630
Val Asp Ser Glu Tyr Tyr Asn Ser Leu Arg Trp Ile Leu Glu Asn
635          640          645
Asp Pro Thr Glu Leu Asp Leu Arg Phe Ile Ile Asp Glu Glu Leu
650          655          660
Phe Gly Gln Thr His Gln His Glu Leu Lys Asn Gly Gly Ser Glu
665          670          675
Ile Val Val Thr Asn Lys Asn Lys Lys Glu Tyr Ile Tyr Leu Val
680          685          690
Ile Gln Trp Arg Phe Val Asn Arg Ile Gln Lys Gln Met Ala Ala
695          700          705
Phe Lys Glu Gly Phe Phe Glu Leu Ile Pro Gln Asp Leu Ile Lys
710          715          720
Ile Phe Asp Glu Asn Glu Leu Glu Leu Leu Met Cys Gly Leu Gly
725          730          735
Asp Val Asp Val Asn Asp Trp Arg Glu His Thr Lys Tyr Lys Asn
740          745          750
Gly Tyr Ser Ala Asn His Gln Val Ile Gln Trp Phe Trp Lys Ala
755          760          765
Val Leu Met Met Asp Ser Glu Lys Arg Ile Arg Leu Leu Gln Phe
770          775          780
Val Thr Gly Thr Ser Arg Val Pro Met Asn Gly Phe Ala Glu Leu
785          790          795
Tyr Gly Ser Asn Gly Pro Gln Ser Phe Thr Val Glu Gln Trp Gly
800          805          810
Thr Pro Glu Lys Leu Pro Arg Ala His Thr Cys Phe Asn Arg Leu
815          820          825
Asp Leu Pro Pro Tyr Glu Ser Phe Glu Glu Leu Trp Asp Lys Leu
830          835          840
Gln Met Ala Ile Glu Asn Thr Gln Gly Phe Asp Gly Val Asp
845          850

```

<210> 4

<211> 111

<212> FRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3016416CD1

<400> 4

```

Met Val Ser Leu Trp Val Glu Asp Thr Phe Leu Ser Pro Gly Phe
1          5          10          15
Gly Phe Ala His Val Ala Cys Ser Gly Leu Gly Met Lys Gln Lys
20          25          30
Arg Lys Ala Ala Ser Ser Glu Pro Thr Ser Glu Val Ala Leu Gly
35          40          45
Gly Ser Ala Gly Pro Val Arg Ser His Leu His Pro Glu Gly Leu
50          55          60
Leu Trp Cys Ser Arg Cys Phe Phe Ser Leu Arg Pro Lys Gly Thr
65          70          75
Glu Pro Pro Gly Arg Ser Ala Gly Leu Gln Gly Ala Thr Glu Arg
80          85          90

```

WO 02/097032

PCT/US02/11152

Ser Gly Trp Thr Ser Val Gln Ala Gln Ala Gln Ala Cys Glu Asn
95 100 105
Leu Val Pro Ala Ala Val
110

<210> 5
<211> 538
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2133755CD1

<400> 5
Met Trp Ser Gly Arg Ser Ser Phe Thr Ser Leu Val Val Gly Val
1 5 10 15
Phe Val Val Tyr Val Val His Thr Cys Trp Val Met Tyr Gly Ile
20 25 30
Val Tyr Thr Arg Pro Cys Ser Gly Asp Ala Asn Cys Ile Gln Pro
35 40 45
Tyr Leu Ala Arg Arg Pro Lys Leu Gln Leu Ser Val Tyr Thr Thr
50 55 60
Thr Arg Ser His Leu Gly Ala Glu Asn Asn Ile Asp Leu Val Leu
65 70 75
Asn Val Glu Asp Phe Asp Val Glu Ser Lys Phe Glu Arg Thr Val
80 85 90
Asn Val Ser Val Pro Lys Lys Thr Arg Asn Asn Gly Thr Leu Tyr
95 100 105
Ala Tyr Ile Phe Leu His His Ala Gly Val Leu Pro Trp His Asp
110 115 120
Gly Lys Gln Val His Leu Val Ser Pro Leu Thr Thr Tyr Met Val
125 130 135
Pro Lys Pro Glu Glu Ile Asn Leu Leu Thr Gly Glu Ser Asp Thr
140 145 150
Gln Gln Ile Glu Ala Glu Lys Lys Pro Thr Ser Ala Leu Asp Glu
155 160 165
Pro Val Ser His Trp Arg Pro Arg Leu Ala Leu Asn Val Met Ala
170 175 180
Asp Asn Phe Val Phe Asp Gly Ser Ser Leu Pro Ala Asp Val His
185 190 195
Arg Tyr Met Lys Met Ile Gln Leu Gly Lys Thr Val His Tyr Leu
200 205 210
Pro Ile Leu Phe Ile Asp Gln Leu Ser Asn Arg Val Lys Asp Leu
215 220 225
Met Val Ile Asn Arg Ser Thr Thr Glu Leu Pro Leu Thr Val Ser
230 235 240
Tyr Asp Lys Val Ser Leu Gly Arg Leu Arg Phe Trp Ile His Met
245 250 255
Gln Asp Ala Val Tyr Ser Leu Gln Gln Phe Gly Phe Ser Glu Lys
260 265 270
Asp Ala Asp Glu Val Lys Gly Ile Phe Val Asp Thr Asn Leu Tyr
275 280 285
Phe Leu Ala Leu Thr Phe Phe Val Ala Ala Phe His Leu Leu Phe
290 295 300
Asp Phe Leu Ala Phe Lys Asn Asp Ile Ser Phe Trp Lys Lys Lys

WO 02/097032

PCT/US02/11152

305 310 315
Lys Ser Met Ile Gly Met Ser Thr Lys Ala Val Leu Trp Arg Cys
320 325 330
Phe Ser Thr Val Val Ile Phe Leu Phe Leu Leu Asp Glu Gln Thr
335 340 345
Ser Leu Leu Val Leu Val Pro Ala Gly Val Gly Ala Ala Ile Glu
350 355 360
Leu Trp Lys Val Lys Lys Ala Leu Lys Met Thr Ile Phe Trp Arg
365 370 375
Gly Leu Met Pro Glu Phe Gln Phe Gly Thr Tyr Ser Glu Ser Glu
380 385 390
Arg Lys Thr Glu Glu Tyr Asp Thr Gln Ala Met Lys Tyr Leu Ser
395 400 405
Tyr Leu Leu Tyr Pro Leu Cys Val Gly Gly Ala Val Tyr Ser Leu
410 415 420
Leu Asn Ile Lys Tyr Lys Ser Trp Tyr Ser Trp Leu Ile Asn Ser
425 430 435
Phe Val Asn Gly Val Tyr Ala Phe Gly Phe Leu Phe Met Leu Pro
440 445 450
Gln Leu Phe Val Asn Tyr Lys Leu Lys Ser Val Ala His Leu Pro
455 460 465
Trp Lys Ala Phe Thr Tyr Lys Ala Phe Asn Thr Phe Ile Asp Asp
470 475 480
Val Phe Ala Phe Ile Ile Thr Met Pro Thr Ser His Arg Leu Ala
485 490 495
Cys Phe Arg Asp Asp Val Val Phe Leu Val Tyr Leu Tyr Gln Arg
500 505 510
Trp Leu Tyr Pro Val Asp Lys Arg Arg Val Asn Glu Phe Gly Glu
515 520 525
Ser Tyr Glu Glu Lys Ala Thr Arg Ala Pro His Thr Asp
530 535

<210> 6

<211> 474

<212> FRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5259957CD1

<400> 6

Met Ser Ile Arg Ala Pro Pro Arg Leu Leu Glu Leu Ala Arg Gln
1 5 10 15
Arg Leu Leu Arg Asp Gln Ala Leu Ala Ile Ser Thr Met Glu Glu
20 25 30
Leu Pro Arg Glu Leu Phe Pro Thr Leu Phe Met Glu Ala Phe Ser
35 40 45
Arg Arg Arg Cys Glu Thr Leu Lys Thr Met Val Gln Ala Trp Pro
50 55 60
Phe Thr Arg Leu Pro Leu Gly Ser Leu Met Lys Ser Pro His Leu
65 70 75
Glu Ser Leu Lys Ser Val Leu Glu Gly Val Asp Val Leu Leu Thr
80 85 90
Gln Glu Val Arg Pro Arg Gln Ser Lys Leu Gln Val Leu Asp Leu
95 100 105

WO 02/097032

PCT/US02/11152

Arg Asn Val Asp Glu Asn Phe Cys Asp Ile Phe Ser Gly Ala Thr
 110 115 120
 Ala Ser Phe Pro Glu Ala Leu Ser Gln Lys Gln Thr Ala Asp Asn
 125 130 135
 Cys Pro Gly Thr Gly Arg Gln Gln Pro Phe Met Val Phe Ile Asp
 140 145 150
 Leu Cys Leu Lys Asn Arg Thr Leu Asp Glu Cys Leu Thr His Leu
 155 160 165
 Leu Glu Trp Gly Lys Gln Arg Lys Gly Leu Leu His Val Cys Cys
 170 175 180
 Lys Glu Leu Gln Val Phe Gly Met Pro Ile His Ser Ile Ile Glu
 185 190 195
 Val Leu Asn Met Val Glu Leu Asp Cys Ile Gln Glu Val Glu Val
 200 205 210
 Cys Cys Pro Trp Glu Leu Ser Thr Leu Val Lys Phe Ala Pro Tyr
 215 220 225
 Leu Gly Gln Met Arg Asn Leu Arg Lys Leu Val Leu Phe Asn Ile
 230 235 240
 Arg Ala Ser Ala Cys Ile Pro Pro Asp Asn Lys Gly Gln Phe Ile
 245 250 255
 Ala Arg Phe Thr Ser Gln Phe Leu Lys Leu Asp Tyr Phe Gln Asn
 260 265 270
 Leu Ser Met His Ser Val Ser Phe Leu Glu Gly His Leu Asp Gln
 275 280 285
 Leu Leu Arg Cys Leu Gln Ala Ser Leu Glu Met Val Val Met Thr
 290 295 300
 Asp Cys Leu Leu Ser Glu Ser Asp Leu Lys His Leu Ser Trp Cys
 305 310 315
 Pro Ser Ile Arg Gln Leu Lys Glu Leu Asp Leu Arg Gly Val Thr
 320 325 330
 Leu Thr His Phe Ser Pro Glu Pro Leu Thr Gly Leu Leu Glu Gln
 335 340 345
 Ala Val Ala Thr Leu Gln Thr Leu Asp Leu Glu Asp Cys Gly Ile
 350 355 360
 Met Asp Ser Gln Leu Ser Ala Ile Leu Pro Val Leu Ser Arg Cys
 365 370 375
 Ser Gln Leu Ser Thr Phe Ser Phe Cys Gly Asn Leu Ile Ser Met
 380 385 390
 Ala Ala Leu Glu Asn Leu Leu Arg His Thr Val Gly Leu Ser Lys
 395 400 405
 Leu Ser Leu Glu Leu Tyr Pro Ala Pro Leu Glu Ser Tyr Asp Thr
 410 415 420
 Gln Gly Ala Leu Cys Trp Gly Arg Phe Ala Glu Leu Gly Ala Glu
 425 430 435
 Leu Met Asn Thr Leu Arg Asp Leu Arg Gln Pro Lys Ile Ile Val
 440 445 450
 Phe Cys Thr Val Pro Cys Pro Arg Cys Gly Ile Arg Ala Ser Tyr
 455 460 465
 Asp Leu Glu Pro Ser His Cys Leu Cys
 470

<210> 7
 <211> 354
 <212> PRT
 <213> Homo sapiens

WO 02/097032

PCT/US02/11152

<220>

<221> misc_feature

<223> Incyte ID No: 55029783CD1

<400> 7

Met Pro Leu Leu Gly Gln Thr Val Arg Ser Ala Ser Ala Arg Thr
1 5 10 15
Arg Arg Trp Ser Arg Arg Ala Ala Gly Asp Arg Pro Gly Ala Pro
20 25 30
Ser Glu Ala Arg Arg Pro Gln Leu Arg Gly Asp His Gly Ile Leu
35 40 45
Val Asp Arg Val Arg Gly His Trp Arg Ile Ala Ala Gly Ser Cys
50 55 60
Ser Thr Cys Trp Cys Pro Ser Ala Leu Cys Ser Ser Thr Asn Gly
65 70 75
Phe Met Cys Thr Thr Gly Phe Pro Asn Met Ser Leu Thr Leu Val
80 85 90
His Phe Val Val Thr Trp Leu Gly Leu Tyr Ile Cys Gln Lys Leu
95 100 105
Asp Ile Phe Ala Pro Lys Ser Leu Pro Pro Ser Arg Leu Leu Leu
110 115 120
Leu Ala Leu Ser Phe Cys Gly Phe Val Val Phe Thr Asn Leu Ser
125 130 135
Leu Gln Asn Asn Thr Ile Gly Thr Tyr Gln Leu Ala Lys Ala Met
140 145 150
Thr Thr Pro Val Ile Ile Ala Ile Gln Thr Phe Cys Tyr Gln Lys
155 160 165
Thr Phe Ser Thr Arg Ile Gln Leu Thr Leu Ile Pro Ile Thr Leu
170 175 180
Gly Val Ile Leu Asn Ser Tyr Tyr Asp Val Lys Phe Asn Phe Leu
185 190 195
Gly Met Val Phe Ala Ala Leu Gly Val Leu Val Thr Ser Leu Tyr
200 205 210
Gln Val Trp Val Gly Ala Lys Gln His Glu Leu Gln Val Asn Ser
215 220 225
Met Gln Leu Leu Tyr Tyr Gln Ala Pro Met Ser Ser Ala Met Leu
230 235 240
Leu Val Ala Val Pro Phe Phe Glu Pro Val Phe Gly Glu Gly Gly
245 250 255
Ile Phe Gly Pro Trp Ser Val Ser Ala Leu Leu Met Val Leu Leu
260 265 270
Ser Gly Val Ile Ala Phe Met Val Asn Leu Ser Ile Tyr Trp Ile
275 280 285
Ile Gly Asn Thr Ser Pro Val Thr Tyr Asn Met Phe Gly His Phe
290 295 300
Lys Phe Cys Ile Thr Leu Phe Gly Gly Tyr Val Leu Phe Lys Asp
305 310 315
Pro Leu Ser Ile Asn Gln Ala Leu Gly Ile Leu Cys Thr Leu Phe
320 325 330
Gly Ile Leu Ala Tyr Thr His Phe Lys Leu Ser Glu Gln Glu Gly
335 340 345
Ser Arg Ser Lys Leu Ala Gln Arg Pro
350

<210> 8

<211> 272

WO 02/097032

PCT/US02/11152

<212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 8032202CD1

<400> 8
 Met Met Cys Pro Leu Trp Arg Leu Leu Ile Phe Leu Gly Leu Leu
 1 5 10 15
 Ala Leu Pro Leu Ala Pro His Lys Gln Pro Trp Pro Gly Leu Ala
 20 25 30
 Gln Ala His Arg Asp Asn Lys Ser Thr Leu Ala Arg Ile Ile Ala
 35 40 45
 Gln Gly Leu Ile Lys His Asn Ala Glu Ser Arg Ile Gln Asn Ile
 50 55 60
 His Phe Gly Asp Arg Leu Asn Ala Ser Ala Gln Val Ala Pro Gly
 65 70 75
 Leu Val Gly Trp Leu Ile Ser Gly Arg Lys His Gln Gln Gln Gln
 80 85 90
 Glu Ser Ser Ile Asn Ile Thr Asn Ile Gln Leu Asp Cys Gly Gly
 95 100 105
 Ile Gln Ile Ser Phe His Lys Glu Trp Phe Ser Ala Asn Ile Ser
 110 115 120
 Leu Glu Phe Asp Leu Glu Leu Arg Pro Ser Phe Asp Asn Asn Ile
 125 130 135
 Val Lys Met Cys Ala His Met Ser Ile Val Val Glu Phe Trp Leu
 140 145 150
 Glu Lys Asp Glu Phe Gly Arg Arg Asp Leu Val Ile Gly Lys Cys
 155 160 165
 Asp Ala Glu Pro Ser Ser Val His Val Ala Ile Leu Thr Glu Ala
 170 175 180
 Ile Pro Pro Lys Met Asn Gln Phe Leu Tyr Asn Leu Lys Glu Asn
 185 190 195
 Leu Gln Lys Val Leu Pro His Met Val Glu Ser Gln Val Cys Pro
 200 205 210
 Leu Ile Gly Glu Ile Leu Gly Gln Leu Asp Val Lys Leu Leu Lys
 215 220 225
 Ser Leu Ile Glu Gln Glu Ala Ala His Glu Pro Thr His His Glu
 230 235 240
 Thr Ser Gln Pro Ser Cys Met Pro Gly Trp Arg Val Pro Gln Leu
 245 250 255
 Thr Ser Ala Asp Gln Lys Glu Ser Pro His Leu Ala Thr Leu Ser
 260 265 270
 Leu Pro

<210> 9
 <211> 710
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 6937367CD1

WO 02/097032

PCT/US02/11152

```

<400> 9
Met Glu Arg Thr Ala Gly Lys Glu Leu Ala Leu Ala Pro Leu Gln
1      5      10      15
Asp Trp Gly Glu Glu Thr Glu Asp Gly Ala Val Tyr Ser Val Ser
20     25     30
Leu Arg Arg Gln Arg Ser Gln Arg Arg Ser Pro Ala Glu Gly Pro
35     40     45
Gly Gly Ser Gln Ala Pro Ser Pro Ile Ala Asn Thr Phe Leu His
50     55     60
Tyr Arg Thr Ser Lys Val Arg Val Leu Arg Ala Ala Arg Leu Glu
65     70     75
Arg Leu Val Gly Glu Leu Val Phe Gly Asp Arg Glu Gln Asp Pro
80     85     90
Ser Phe Met Pro Ala Phe Leu Ala Thr Tyr Arg Thr Phe Val Pro
95     100    105
Thr Ala Cys Leu Leu Gly Phe Leu Leu Pro Pro Met Pro Pro Pro
110    115    120
Pro Pro Pro Gly Val Glu Ile Lys Lys Thr Ala Val Gln Asp Leu
125    130    135
Ser Phe Asn Lys Asn Leu Arg Ala Val Val Ser Val Leu Gly Ser
140    145    150
Trp Leu Gln Asp His Pro Gln Asp Phe Arg Asp His Pro Ala His
155    160    165
Ser Asp Leu Gly Ser Val Arg Thr Phe Leu Gly Trp Ala Ala Pro
170    175    180
Gly Ser Ala Glu Ala Gln Lys Ala Glu Lys Leu Leu Glu Asp Phe
185    190    195
Leu Glu Glu Ala Glu Arg Glu Gln Glu Glu Glu Pro Pro Gln Val
200    205    210
Trp Thr Gly Pro Pro Arg Val Ala Gln Thr Ser Asp Pro Asp Ser
215    220    225
Ser Glu Ala Cys Ala Glu Glu Glu Glu Gly Leu Met Pro Gln Gly
230    235    240
Pro Gln Leu Leu Asp Phe Ser Val Asp Glu Val Ala Glu Gln Leu
245    250    255
Thr Leu Ile Asp Leu Glu Leu Phe Ser Lys Val Arg Leu Tyr Glu
260    265    270
Cys Leu Gly Ser Val Trp Ser Gln Arg Asp Arg Pro Gly Ala Ala
275    280    285
Gly Ala Ser Pro Thr Val Arg Ala Thr Val Ala Gln Phe Asn Thr
290    295    300
Val Thr Gly Cys Val Leu Gly Ser Val Leu Gly Ala Pro Gly Leu
305    310    315
Ala Ala Pro Gln Arg Ala Gln Arg Leu Glu Lys Trp Ile Arg Ile
320    325    330
Ala Gln Arg Cys Arg Glu Leu Arg Asn Phe Ser Ser Leu Arg Ala
335    340    345
Ile Leu Ser Ala Leu Gln Ser Asn Pro Ile Tyr Arg Leu Lys Arg
350    355    360
Ser Trp Gly Ala Val Ser Arg Glu Pro Leu Ser Thr Phe Arg Lys
365    370    375
Leu Ser Gln Ile Phe Ser Asp Glu Asn Asn His Leu Ser Ser Arg
380    385    390
Glu Ile Leu Phe Gln Glu Glu Ala Thr Glu Gly Ser Gln Glu Glu
395    400    405
Asp Asn Thr Pro Gly Ser Leu Pro Ser Lys Pro Pro Pro Gly Pro

```

WO 02/097032

PCT/US02/11152

410	415	420
Val Pro Tyr Leu Gly Thr Phe Leu Thr	Asp Leu Val Met Leu Asp	
425	430	435
Thr Ala Leu Pro Asp Met Leu Glu Gly	Asp Leu Ile Asn Phe Glu	
440	445	450
Lys Arg Arg Lys Glu Trp Glu Ile Leu	Ala Arg Ile Gln Gln Leu	
455	460	465
Gln Arg Arg Cys Gln Ser Tyr Thr Leu	Ser Pro His Pro Pro Ile	
470	475	480
Leu Ala Ala Leu His Ala Gln Asn Gln	Leu Thr Glu Glu Gln Ser	
485	490	495
Tyr Arg Leu Ser Arg Val Ile Glu Pro	Pro Ala Ala Ser Cys Pro	
500	505	510
Ser Ser Pro Arg Ile Arg Arg Arg Ile	Ser Leu Thr Lys Arg Leu	
515	520	525
Ser Ala Lys Leu Ala Arg Glu Lys Ser	Ser Ser Pro Ser Gly Ser	
530	535	540
Pro Gly Asp Pro Ser Ser Pro Thr Ser	Ser Val Ser Pro Gly Ser	
545	550	555
Pro Pro Ser Ser Pro Arg Ser Arg Asp	Ala Pro Ala Gly Ser Pro	
560	565	570
Pro Ala Ser Pro Gly Pro Gln Gly Pro	Ser Thr Lys Leu Pro Leu	
575	580	585
Ser Leu Asp Leu Pro Ser Pro Arg Pro	Phe Ala Leu Pro Leu Gly	
590	595	600
Ser Pro Arg Ile Pro Leu Pro Ala Gln	Gln Ser Ser Glu Ala Arg	
605	610	615
Val Ile Arg Val Ser Ile Asp Asn Asp	His Gly Asn Leu Tyr Arg	
620	625	630
Ser Ile Leu Leu Thr Ser Gln Asp Lys	Ala Pro Ser Val Val Arg	
635	640	645
Arg Ala Leu Gln Lys His Asn Val Pro	Gln Pro Trp Ala Cys Asp	
650	655	660
Tyr Gln Leu Phe Gln Val Leu Pro Gly	Asp Arg Val Leu Leu Ile	
665	670	675
Pro Asp Asn Ala Asn Val Phe Tyr Ala	Met Ser Pro Val Ala Pro	
680	685	690
Arg Asp Phe Met Leu Arg Arg Lys Glu	Gly Thr Arg Asn Thr Leu	
695	700	705
Ser Val Ser Pro Ser		
710		

<210> 10

<211> 490

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3876510CD1

<400> 10

Met Thr Ile Gly Arg Met Glu Asn Val	Glu Val Phe Thr Ala Glu		
1	5	10	15
Gly Lys Gly Arg Gly Leu Lys Ala Thr	Lys Glu Phe Trp Ala Ala		
20	25	30	

WO 02/097032

PCT/US02/11152

```

Asp Ile Ile Phe Ala Glu Arg Ala Tyr Ser Ala Val Val Phe Asp
35 40 45
Ser Leu Val Asn Phe Val Cys His Thr Cys Phe Lys Arg Gln Glu
50 55 60
Lys Leu His Arg Cys Gly Gln Cys Lys Phe Ala His Tyr Cys Asp
65 70 75
Arg Thr Cys Gln Lys Asp Ala Trp Leu Asn His Lys Asn Glu Cys
80 85 90
Ser Ala Ile Lys Arg Tyr Gly Lys Val Pro Asn Glu Asn Ile Arg
95 100 105
Leu Ala Ala Arg Ile Met Trp Arg Val Glu Arg Glu Gly Thr Gly
110 115 120
Leu Thr Glu Gly Cys Leu Val Ser Val Asp Asp Leu Gln Asn His
125 130 135
Val Glu His Phe Gly Glu Glu Glu Gln Lys Asp Leu Arg Val Asp
140 145 150
Val Asp Thr Phe Leu Gln Tyr Trp Pro Pro Gln Ser Gln Gln Phe
155 160 165
Ser Met Gln Tyr Ile Ser His Ile Phe Gly Val Ile Asn Cys Asn
170 175 180
Gly Phe Thr Leu Ser Asp Gln Arg Gly Leu Gln Ala Val Gly Val
185 190 195
Gly Ile Phe Pro Asn Leu Gly Leu Val Asn His Asp Cys Trp Pro
200 205 210
Asn Cys Thr Val Ile Phe Asn Asn Gly Asn His Glu Ala Val Lys
215 220 225
Ser Met Phe His Thr Gln Met Arg Ile Glu Leu Arg Ala Leu Gly
230 235 240
Lys Ile Ser Glu Gly Glu Glu Leu Thr Val Ser Tyr Ile Asp Phe
245 250 255
Leu Asn Val Ser Glu Glu Arg Lys Arg Gln Leu Lys Lys Gln Tyr
260 265 270
Tyr Phe Asp Cys Thr Cys Glu His Cys Gln Lys Lys Leu Lys Asp
275 280 285
Asp Leu Phe Leu Gly Val Lys Asp Asn Pro Lys Pro Ser Gln Glu
290 295 300
Val Val Lys Glu Met Ile Gln Phe Ser Lys Asp Thr Leu Glu Lys
305 310 315
Ile Asp Lys Ala Arg Ser Glu Gly Leu Tyr His Glu Val Val Lys
320 325 330
Leu Cys Arg Glu Cys Leu Glu Lys Gln Glu Pro Val Phe Ala Asp
335 340 345
Thr Asn Ile Tyr Met Leu Arg Met Leu Ser Ile Val Ser Glu Val
350 355 360
Leu Ser Tyr Leu Gln Ala Phe Glu Glu Ala Ser Phe Tyr Ala Arg
365 370 375
Arg Met Val Asp Gly Tyr Met Lys Leu Tyr His Pro Asn Asn Ala
380 385 390
Gln Leu Gly Met Ala Val Met Arg Ala Gly Leu Thr Asn Trp His
395 400 405
Ala Gly Asn Ile Glu Val Gly His Gly Met Ile Cys Lys Ala Tyr
410 415 420
Ala Ile Leu Leu Val Thr His Gly Pro Ser His Pro Ile Thr Lys
425 430 435
Asp Leu Glu Ala Met Arg Val Gln Thr Glu Met Glu Leu Arg Met
440 445 450

```

WO 02/097032

PCT/US02/11152

Phe Arg Gln Asn Glu Phe Met Tyr Tyr Lys Met Arg Glu Ala Ala
455 460 465
Leu Asn Asn Gln Pro Met Gln Val Met Ala Glu Pro Ser Asn Glu
470 475 480
Pro Ser Pro Ala Leu Phe His Lys Lys Gln
485 490

<210> 11
<211> 599
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4900076CD1

<400> 11
Met Met Leu Pro Tyr Pro Ser Ala Leu Gly Asp Gln Tyr Trp Glu
1 5 10 15
Glu Ile Leu Leu Pro Lys Asn Gly Glu Asn Val Glu Thr Met Lys
20 25 30
Lys Leu Thr Gln Asn His Lys Ala Lys Gly Leu Pro Ser Asn Asp
35 40 45
Thr Asp Cys Pro Gln Lys Lys Glu Gly Lys Ala Gln Ile Val Val
50 55 60
Pro Val Thr Phe Arg Asp Val Thr Val Ile Phe Thr Glu Ala Glu
65 70 75
Trp Lys Arg Leu Ser Pro Glu Gln Arg Asn Leu Tyr Lys Glu Val
80 85 90
Met Leu Glu Asn Tyr Arg Asn Leu Leu Ser Leu Ala Glu Pro Lys
95 100 105
Pro Glu Ile Tyr Thr Cys Ser Ser Cys Leu Leu Ala Phe Ser Cys
110 115 120
Gln Gln Phe Leu Ser Gln His Val Leu Gln Ile Phe Leu Gly Leu
125 130 135
Cys Ala Glu Asn His Phe His Pro Gly Asn Ser Ser Pro Gly His
140 145 150
Trp Lys Gln Gln Gly Gln Gln Tyr Ser His Val Ser Cys Trp Phe
155 160 165
Glu Asn Ala Glu Gly Gln Glu Arg Gly Gly Gly Ser Lys Pro Trp
170 175 180
Ser Ala Arg Thr Glu Glu Arg Glu Thr Ser Arg Ala Phe Pro Ser
185 190 195
Pro Leu Gln Arg Gln Ser Ala Ser Pro Arg Lys Gly Asn Met Val
200 205 210
Val Glu Thr Glu Pro Ser Ser Ala Gln Arg Pro Asn Pro Val Gln
215 220 225
Leu Asp Lys Gly Leu Lys Glu Leu Glu Thr Leu Arg Phe Gly Ala
230 235 240
Ile Asn Cys Arg Glu Tyr Glu Pro Asp His Asn Leu Glu Ser Asn
245 250 255
Phe Ile Thr Asn Pro Arg Thr Leu Leu Gly Lys Lys Pro Tyr Ile
260 265 270
Cys Ser Asp Cys Gly Arg Ser Phe Lys Asp Arg Ser Thr Leu Ile
275 280 285
Arg His His Arg Ile His Ser Met Glu Lys Pro Tyr Val Cys Ser

WO 02/097032

PCT/US02/11152

```

      290      295      300
Glu Cys Gly Arg Gly Phe Ser Gln Lys Ser Asn Leu Ser Arg His
      305      310      315
Gln Arg Thr His Ser Glu Glu Lys Pro Tyr Leu Cys Arg Glu Cys
      320      325      330
Gly Gln Ser Phe Arg Ser Lys Ser Ile Leu Asn Arg His Gln Trp
      335      340      345
Thr His Ser Glu Glu Lys Pro Tyr Val Cys Ser Glu Cys Gly Arg
      350      355      360
Gly Phe Ser Glu Lys Ser Ser Phe Ile Arg His Gln Arg Thr His
      365      370      375
Ser Gly Glu Lys Pro Tyr Val Cys Leu Glu Cys Gly Arg Ser Phe
      380      385      390
Cys Asp Lys Ser Thr Leu Arg Lys His Gln Arg Ile His Ser Gly
      395      400      405
Glu Lys Pro Tyr Val Cys Arg Glu Cys Gly Arg Gly Phe Ser Gln
      410      415      420
Asn Ser Asp Leu Ile Lys His Gln Arg Thr His Leu Asp Glu Lys
      425      430      435
Pro Tyr Val Cys Arg Glu Cys Gly Arg Gly Phe Cys Asp Lys Ser
      440      445      450
Thr Leu Ile Ile His Glu Arg Thr His Ser Gly Glu Lys Pro Tyr
      455      460      465
Val Cys Gly Glu Cys Gly Arg Gly Phe Ser Arg Lys Ser Leu Leu
      470      475      480
Leu Val His Gln Arg Thr His Ser Gly Glu Lys His Tyr Val Cys
      485      490      495
Arg Glu Cys Arg Arg Gly Phe Ser Gln Lys Ser Asn Leu Ile Arg
      500      505      510
His Gln Arg Thr His Ser Asn Glu Lys Pro Tyr Ile Cys Arg Glu
      515      520      525
Cys Gly Arg Gly Phe Cys Asp Lys Ser Thr Leu Ile Val His Glu
      530      535      540
Arg Thr His Ser Gly Glu Lys Pro Tyr Val Cys Ser Glu Cys Gly
      545      550      555
Arg Gly Phe Ser Arg Lys Ser Leu Leu Leu Val His Gln Arg Thr
      560      565      570
His Ser Gly Glu Lys His Tyr Val Cys Arg Glu Cys Gly Arg Gly
      575      580      585
Phe Ser His Lys Ser Asn Leu Ile Arg His Gln Arg Thr His
      590      595

```

<210> 12

<211> 365

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1543848CD1

<400> 12

```

Met Ala Ala Ala Ala Ala Gly Thr Ala Thr Ser Gln Arg Phe Phe
  1           5           10          15
Gln Ser Phe Ser Asp Ala Leu Ile Asp Glu Asp Pro Gln Ala Ala
      20           25           30

```


WO 02/097032

PCT/US02/11152

```

Leu Glu Glu Leu Thr Lys Ala Leu Glu Gln Lys Pro Asp Asp Ala
   35                               40           45
Gln Tyr Tyr Cys Gln Arg Ala Tyr Cys His Ile Leu Leu Gly Asn
   50                               55           60
Tyr Cys Val Ala Val Ala Asp Ala Lys Lys Ser Leu Glu Leu Asn
   65                               70           75
Pro Asn Asn Ser Thr Ala Met Leu Arg Lys Gly Ile Cys Glu Tyr
   80                               85           90
His Glu Lys Asn Tyr Ala Ala Ala Leu Glu Thr Phe Thr Glu Gly
   95                               100          105
Gln Lys Leu Asp Ile Glu Thr Gly Phe His Arg Val Gly Gln Ala
  110                               115          120
Gly Leu Gln Leu Leu Thr Ser Ser Asp Pro Pro Ala Leu Asp Ser
  125                               130          135
Gln Ser Ala Gly Ile Thr Gly Ala Asp Ala Asn Phe Ser Val Trp
  140                               145          150
Ile Lys Arg Cys Gln Glu Ala Gln Asn Gly Ser Glu Ser Glu Val
  155                               160          165
Trp Thr His Gln Ser Lys Ile Lys Tyr Asp Trp Tyr Gln Thr Glu
  170                               175          180
Ser Gln Val Val Ile Thr Leu Met Ile Lys Asn Val Gln Lys Asn
  185                               190          195
Asp Val Asn Val Glu Phe Ser Glu Lys Glu Leu Ser Ala Leu Val
  200                               205          210
Lys Leu Pro Ser Gly Glu Asp Tyr Asn Leu Lys Leu Glu Leu Leu
  215                               220          225
His Pro Ile Ile Pro Glu Gln Ser Thr Phe Lys Val Leu Ser Thr
  230                               235          240
Lys Ile Glu Ile Lys Leu Lys Lys Pro Glu Ala Val Arg Trp Glu
  245                               250          255
Lys Leu Glu Gly Gln Gly Asp Val Pro Thr Pro Lys Gln Phe Val
  260                               265          270
Ala Asp Val Lys Asn Leu Tyr Pro Ser Ser Ser Pro Tyr Thr Arg
  275                               280          285
Asn Trp Asp Lys Leu Val Gly Glu Ile Lys Glu Glu Glu Lys Asn
  290                               295          300
Glu Lys Leu Glu Gly Asp Ala Ala Leu Asn Arg Leu Phe Gln Gln
  305                               310          315
Ile Tyr Ser Asp Gly Ser Asp Glu Val Lys Arg Ala Met Asn Lys
  320                               325          330
Ser Phe Met Glu Ser Gly Gly Thr Val Leu Ser Thr Asn Trp Ser
  335                               340          345
Asp Val Gly Lys Arg Lys Val Glu Ile Asn Pro Pro Asp Asp Met
  350                               355          360
Glu Trp Lys Lys Tyr
  365

```

<210> 13
 <211> 365
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 6254070CD1

WO 02/097032

PCT/US02/11152

```

<400> 13
Met Ala Gly Ser Ala Met Ser Ser Lys Phe Phe Leu Val Ala Leu
1      5      10      15
Ala Ile Phe Phe Ser Phe Ala Gln Val Val Ile Glu Ala Asn Ser
20     25     30
Trp Trp Ser Leu Gly Met Asn Asn Pro Val Gln Met Ser Glu Val
35     40     45
Tyr Ile Ile Gly Ala Gln Pro Leu Cys Ser Gln Leu Ala Gly Leu
50     55     60
Ser Gln Gly Gln Lys Lys Leu Cys His Leu Tyr Gln Asp His Met
65     70     75
Gln Tyr Ile Gly Glu Gly Ala Lys Thr Gly Ile Lys Glu Cys Gln
80     85     90
Tyr Gln Phe Arg His Arg Arg Trp Asn Cys Ser Thr Val Asp Asn
95     100    105
Thr Ser Val Phe Gly Arg Val Met Gln Ile Gly Ser Arg Glu Thr
110    115    120
Ala Phe Thr Tyr Ala Val Ser Ala Ala Gly Val Val Asn Ala Met
125    130    135
Ser Arg Ala Cys Arg Glu Gly Glu Leu Ser Thr Cys Gly Cys Ser
140    145    150
Arg Ala Ala Arg Pro Lys Asp Leu Pro Arg Asp Trp Leu Trp Gly
155    160    165
Gly Cys Gly Asp Asn Ile Asp Tyr Gly Tyr Arg Phe Ala Lys Glu
170    175    180
Phe Val Asp Ala Arg Glu Arg Glu Arg Ile His Ala Lys Gly Ser
185    190    195
Tyr Glu Ser Ala Arg Ile Leu Met Asn Leu His Asn Asn Glu Ala
200    205    210
Gly Arg Arg Thr Val Tyr Asn Leu Ala Asp Val Ala Cys Lys Cys
215    220    225
His Gly Val Ser Gly Ser Cys Ser Leu Lys Thr Cys Trp Leu Gln
230    235    240
Leu Ala Asp Phe Arg Lys Val Gly Asp Ala Leu Lys Glu Lys Tyr
245    250    255
Asp Ser Ala Ala Ala Met Arg Leu Asn Ser Arg Gly Lys Leu Val
260    265    270
Gln Val Asn Ser Arg Phe Asn Ser Pro Thr Thr Gln Asp Leu Val
275    280    285
Tyr Ile Asp Pro Ser Pro Asp Tyr Cys Val Arg Asn Glu Ser Thr
290    295    300
Gly Ser Leu Gly Thr Gln Gly Arg Leu Cys Asn Lys Thr Ser Glu
305    310    315
Gly Met Asp Gly Cys Glu Leu Met Cys Cys Gly Arg Gly Tyr Asp
320    325    330
Gln Phe Lys Thr Val Gln Thr Glu Arg Cys His Cys Lys Phe His
335    340    345
Trp Cys Cys Tyr Val Lys Cys Lys Lys Cys Thr Glu Ile Val Asp
350    355    360
Gln Phe Val Cys Lys
365

```

```

<210> 14
<211> 203
<212> PRT
<213> Homo sapiens

```

WO 02/097032

PCT/US02/11152

<220>

<221> misc_feature

<223> Incyte ID No: 1289839CD1

<400> 14

```

Met Val Ser Thr Tyr Arg Val Ala Val Leu Gly Ala Arg Gly Val
 1          5          10          15
Gly Lys Ser Ala Ile Val Arg Gln Phe Leu Tyr Asn Glu Phe Ser
 20          25          30
Glu Val Cys Val Pro Thr Thr Ala Arg Arg Leu Tyr Leu Pro Ala
 35          40          45
Val Val Met Asn Gly His Val His Asp Leu Gln Ile Leu Asp Phe
 50          55          60
Pro Pro Ile Ser Ala Phe Pro Val Asn Thr Leu Gln Glu Trp Ala
 65          70          75
Asp Thr Cys Cys Arg Gly Leu Arg Ser Val His Ala Tyr Ile Leu
 80          85          90
Val Tyr Asp Ile Cys Cys Phe Asp Ser Phe Glu Tyr Val Lys Thr
 95          100          105
Ile Arg Gln Gln Ile Leu Glu Thr Arg Val Ile Gly Thr Ser Glu
 110          115          120
Thr Pro Ile Ile Ile Val Gly Asn Lys Arg Asp Leu Gln Arg Gly
 125          130          135
Arg Val Ile Pro Arg Trp Asn Val Ser His Leu Val Arg Lys Thr
 140          145          150
Trp Lys Cys Gly Tyr Val Glu Cys Ser Ala Lys Tyr Asn Trp His
 155          160          165
Ile Leu Leu Leu Phe Ser Glu Leu Leu Lys Ser Val Gly Cys Ala
 170          175          180
Arg Cys Lys His Val His Ala Ala Leu Arg Phe Gln Gly Ala Leu
 185          190          195
Arg Arg Asn Arg Cys Ala Ile Met
 200

```

<210> 15

<211> 403

<212> FRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5565648CD1

<400> 15

```

Met Glu Pro Val Gly Cys Cys Gly Glu Cys Arg Gly Ser Ser Val
 1          5          10          15
Asp Pro Arg Ser Thr Phe Val Leu Ser Asn Leu Ala Glu Val Val
 20          25          30
Glu Arg Val Leu Thr Phe Leu Pro Ala Lys Ala Leu Leu Arg Val
 35          40          45
Ala Cys Val Cys Arg Leu Trp Arg Glu Cys Val Arg Arg Val Leu
 50          55          60
Arg Thr His Arg Ser Val Thr Trp Ile Ser Ala Gly Leu Ala Glu
 65          70          75
Ala Gly His Leu Glu Gly His Cys Leu Val Arg Val Val Ala Glu
 80          85          90

```

WO 02/097032

PCT/US02/11152

Glu Leu Glu Asn Val Arg Ile Leu Pro His Thr Val Leu Tyr Met
 95 100 105
 Ala Asp Ser Glu Thr Phe Ile Ser Leu Glu Glu Cys Arg Gly His
 110 115 120
 Lys Arg Ala Arg Lys Arg Thr Ser Met Glu Thr Ala Leu Ala Leu
 125 130 135
 Glu Lys Leu Phe Pro Lys Gln Cys Gln Val Leu Gly Ile Val Thr
 140 145 150
 Pro Gly Ile Val Val Thr Pro Met Gly Ser Gly Ser Asn Arg Pro
 155 160 165
 Gln Glu Ile Glu Ile Gly Glu Ser Gly Phe Ala Leu Leu Phe Pro
 170 175 180
 Gln Ile Glu Gly Ile Lys Ile Gln Pro Phe His Phe Ile Lys Asp
 185 190 195
 Pro Lys Asn Leu Thr Leu Glu Arg His Gln Leu Thr Glu Val Gly
 200 205 210
 Leu Leu Asp Asn Pro Glu Leu Arg Val Val Leu Val Phe Gly Tyr
 215 220 225
 Asn Cys Cys Lys Val Gly Ala Ser Asn Tyr Leu Gln Gln Val Val
 230 235 240
 Ser Thr Phe Ser Asp Met Asn Ile Ile Leu Ala Gly Gly Gln Val
 245 250 255
 Asp Asn Leu Ser Ser Leu Thr Ser Glu Lys Asn Pro Leu Asp Ile
 260 265 270
 Asp Ala Ser Gly Val Val Gly Leu Ser Phe Ser Gly His Arg Ile
 275 280 285
 Gln Ser Ala Thr Val Leu Leu Asn Glu Asp Val Ser Asp Glu Lys
 290 295 300
 Thr Ala Glu Ala Ala Met Gln Arg Leu Lys Ala Ala Asn Ile Pro
 305 310 315
 Glu His Asn Thr Ile Gly Phe Met Phe Ala Cys Val Gly Arg Gly
 320 325 330
 Phe Gln Tyr Tyr Arg Ala Lys Gly Asn Val Glu Ala Asp Ala Phe
 335 340 345
 Arg Lys Phe Phe Pro Ser Val Pro Leu Phe Gly Phe Phe Gly Asn
 350 355 360
 Gly Glu Ile Gly Cys Asp Arg Ile Val Thr Gly Asn Phe Ile Leu
 365 370 375
 Arg Lys Cys Asn Glu Val Lys Asp Asp Asp Leu Phe His Ser Tyr
 380 385 390
 Thr Thr Ile Met Ala Leu Ile His Leu Gly Ser Ser Lys
 395 400

<210> 16
 <211> 1022
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2754456CD1

<400> 16
 Met Tyr Phe Cys Trp Gly Ala Asp Ser Arg Glu Leu Gln Arg Arg
 1 5 10 15
 Arg Thr Ala Gly Ser Pro Gly Ala Glu Leu Leu Gln Ala Ala Ser

WO 02/097032

PCT/US02/11152

20 25 30
 Gly Glu Arg His Ser Leu Leu Leu Leu Thr Asn His Arg Val Leu
 35 40 45
 Ser Cys Gly Asp Asn Ser Arg Gly Gln Leu Gly Arg Arg Gly Ala
 50 55 60
 Gln Arg Gly Glu Leu Pro Glu Pro Ile Gln Ala Leu Glu Thr Leu
 65 70 75
 Ile Val Asp Leu Val Ser Cys Gly Lys Glu His Ser Leu Ala Val
 80 85 90
 Cys His Lys Gly Arg Val Phe Ala Trp Gly Ala Gly Ser Glu Gly
 95 100 105
 Gln Leu Gly Ile Gly Glu Phe Lys Glu Ile Ser Phe Thr Pro Lys
 110 115 120
 Lys Ile Met Thr Leu Asn Asp Ile Lys Ile Ile Gln Val Ser Cys
 125 130 135
 Gly His Tyr His Ser Leu Ala Leu Ser Lys Asp Ser Gln Val Phe
 140 145 150
 Ser Trp Gly Lys Asn Ser His Gly Gln Leu Gly Leu Gly Lys Glu
 155 160 165
 Phe Pro Ser Gln Ala Ser Pro Gln Arg Val Arg Ser Leu Glu Gly
 170 175 180
 Ile Pro Leu Ala Gln Val Ala Ala Gly Gly Ala His Ser Phe Ala
 185 190 195
 Leu Ser Leu Cys Gly Thr Ser Phe Gly Trp Gly Ser Asn Ser Ala
 200 205 210
 Gly Gln Leu Ala Leu Ser Gly Arg Asn Val Pro Val Gln Ser Asn
 215 220 225
 Lys Pro Leu Ser Val Gly Ala Leu Lys Asn Leu Gly Val Val Tyr
 230 235 240
 Ile Ser Cys Gly Asp Ala His Thr Ala Val Leu Thr Gln Asp Gly
 245 250 255
 Lys Val Phe Thr Phe Gly Asp Asn Arg Ser Gly Gln Leu Gly Tyr
 260 265 270
 Ser Pro Thr Pro Glu Lys Arg Gly Pro Gln Leu Val Glu Arg Ile
 275 280 285
 Asp Gly Leu Val Ser Gln Ile Asp Cys Gly Ser Tyr His Thr Leu
 290 295 300
 Ala Tyr Val His Thr Thr Gly Gln Val Val Ser Phe Gly His Gly
 305 310 315
 Pro Ser Asp Thr Ser Lys Pro Thr His Pro Glu Ala Leu Thr Glu
 320 325 330
 Asn Phe Asp Ile Ser Cys Leu Ile Ser Ala Glu Asp Phe Val Asp
 335 340 345
 Val Gln Val Lys His Ile Phe Ala Gly Thr Tyr Ala Asn Phe Val
 350 355 360
 Thr Thr His Gln Asp Thr Ser Ser Thr Arg Ala Pro Gly Lys Thr
 365 370 375
 Leu Pro Glu Ile Ser Arg Ile Ser Gln Ser Met Ala Glu Lys Trp
 380 385 390
 Ile Ala Val Lys Arg Arg Ser Thr Glu His Glu Met Ala Lys Ser
 395 400 405
 Glu Ile Arg Met Ile Phe Ser Ser Pro Ala Cys Leu Thr Ala Ser
 410 415 420
 Phe Leu Lys Lys Arg Gly Thr Gly Glu Thr Thr Ser Ile Asp Val
 425 430 435
 Asp Leu Glu Met Ala Arg Asp Thr Phe Lys Lys Leu Thr Lys Lys

WO 02/097032

PCT/US02/11152

```

440          445          450
Glu Trp Ile Ser Ser Met Ile Thr Thr Cys Leu Glu Asp Asp Leu
455          460          465
Leu Arg Ala Leu Pro Cys His Ser Pro His Gln Glu Ala Leu Ser
470          475          480
Val Phe Leu Leu Leu Pro Glu Cys Pro Val Met His Asp Ser Lys
485          490          495
Asn Trp Lys Asn Leu Val Val Pro Phe Ala Lys Ala Val Cys Glu
500          505          510
Met Ser Lys Gln Ser Leu Gln Val Leu Lys Lys Cys Trp Ala Phe
515          520          525
Leu Gln Glu Ser Ser Leu Asn Pro Leu Ile Gln Met Leu Lys Ala
530          535          540
Ala Ile Ile Ser Gln Leu Leu His Gln Thr Lys Thr Glu Gln Asp
545          550          555
His Cys Asn Val Lys Ala Leu Leu Gly Met Met Lys Glu Leu His
560          565          570
Lys Val Asn Lys Ala Asn Cys Arg Leu Pro Glu Asn Thr Phe Asn
575          580          585
Ile Asn Glu Leu Ser Asn Leu Leu Asn Phe Tyr Ile Asp Arg Gly
590          595          600
Arg Gln Leu Phe Arg Asp Asn His Leu Ile Pro Ala Glu Thr Pro
605          610          615
Ser Pro Val Ile Phe Ser Asp Phe Pro Phe Ile Phe Asn Ser Leu
620          625          630
Ser Lys Ile Lys Lys Leu Gln Ala Asp Ser His Ile Lys Met Gln
635          640          645
Met Ser Glu Lys Lys Ala Tyr Met Leu Met His Glu Thr Ile Leu
650          655          660
Gln Lys Lys Asp Glu Phe Pro Pro Ser Pro Arg Phe Ile Leu Arg
665          670          675
Val Arg Arg Ser Arg Leu Val Lys Asp Ala Leu Arg Gln Leu Ser
680          685          690
Gln Ala Glu Ala Thr Asp Phe Cys Lys Val Leu Val Val Glu Phe
695          700          705
Ile Asn Glu Ile Cys Pro Glu Ser Gly Gly Val Ser Ser Glu Phe
710          715          720
Phe His Cys Met Phe Glu Glu Met Thr Lys Pro Glu Tyr Gly Met
725          730          735
Phe Met Tyr Pro Glu Met Gly Ser Cys Met Trp Phe Pro Ala Lys
740          745          750
Pro Lys Pro Glu Lys Lys Arg Tyr Phe Leu Phe Gly Met Leu Cys
755          760          765
Gly Leu Ser Leu Phe Asn Leu Asn Val Ala Asn Leu Pro Phe Pro
770          775          780
Leu Ala Leu Tyr Lys Lys Leu Leu Asp Gln Lys Pro Ser Leu Glu
785          790          795
Asp Leu Lys Glu Leu Ser Pro Arg Leu Gly Lys Ser Leu Gln Glu
800          805          810
Val Leu Asp Asp Ala Ala Asp Asp Ile Gly Asp Ala Leu Cys Ile
815          820          825
Arg Phe Ser Ile His Trp Asp Gln Asn Asp Val Asp Leu Ile Pro
830          835          840
Asn Gly Ile Ser Ile Pro Val Asp Gln Thr Asn Lys Arg Asp Tyr
845          850          855
Val Ser Lys Tyr Ile Asp Tyr Ile Phe Asn Val Ser Val Lys Ala

```

WO 02/097032

PCT/US02/11152

```

      860      865      870
Val Tyr Glu Glu Phe Gln Arg Gly Phe Tyr Arg Val Cys Glu Lys
      875      880      885
Glu Ile Leu Arg His Phe Tyr Pro Glu Glu Leu Met Thr Ala Ile
      890      895      900
Ile Gly Asn Thr Asp Tyr Asp Trp Lys Gln Phe Glu Gln Asn Ser
      905      910      915
Lys Tyr Glu Gln Gly Tyr Gln Lys Ser His Pro Thr Ile Gln Leu
      920      925      930
Phe Trp Lys Ala Phe His Lys Leu Thr Leu Asp Glu Lys Lys Lys
      935      940      945
Phe Leu Phe Phe Leu Thr Gly Arg Asp Arg Leu His Ala Arg Gly
      950      955      960
Ile Gln Lys Met Glu Ile Val Phe Arg Cys Pro Glu Thr Phe Ser
      965      970      975
Glu Arg Asp His Pro Thr Ser Ile Thr Cys His Asn Ile Leu Ser
      980      985      990
Leu Pro Lys Tyr Ser Thr Met Glu Arg Met Glu Glu Ala Leu Gln
      995      1000      1005
Val Ala Ile Asn Asn Asn Arg Gly Phe Val Ser Pro Met Leu Thr
      1010      1015      1020
Gln Ser

```

```

<210> 17
<211> 1462
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 5734806CD1

```

```

<400> 17
Met Gly Ala Gln Asp Arg Pro Gln Cys His Phe Asp Ile Glu Ile
      1      5      10      15
Asn Arg Glu Pro Val Gly Arg Ile Met Phe Gln Leu Phe Ser Asp
      20      25      30
Ile Cys Pro Lys Thr Cys Lys Asn Phe Leu Cys Leu Cys Ser Gly
      35      40      45
Glu Lys Gly Leu Gly Lys Thr Thr Gly Lys Lys Leu Cys Tyr Lys
      50      55      60
Gly Ser Thr Phe His Arg Val Val Lys Asn Phe Met Ile Gln Gly
      65      70      75
Gly Asp Phe Ser Glu Gly Asn Gly Lys Gly Glu Ser Ile Tyr
      80      85      90
Gly Gly Tyr Phe Lys Asp Glu Asn Phe Ile Leu Lys His Asp Arg
      95      100      105
Ala Phe Leu Leu Ser Met Ala Asn Arg Gly Lys His Thr Asn Gly
      110      115      120
Ser Gln Phe Phe Ile Thr Thr Lys Pro Ala Pro His Leu Asp Gly
      125      130      135
Val His Val Val Phe Gly Leu Val Ile Ser Gly Phe Glu Val Ile
      140      145      150
Glu Gln Ile Glu Asn Leu Lys Thr Asp Ala Ala Ser Arg Pro Tyr
      155      160      165

```

WO 02/097032

PCT/US02/11152

Ala Asp Val Arg Val Ile Asp Cys Gly Val Leu Ala Thr Lys Ser
 170 175 180
 Ile Lys Asp Val Phe Glu Lys Lys Arg Lys Lys Pro Thr His Ser
 185 190 195
 Glu Gly Ser Asp Ser Ser Ser Asn Ser Ser Ser Ser Ser Glu Ser
 200 205 210
 Ser Ser Glu Ser Glu Leu Glu His Glu Arg Ser Arg Arg Arg Lys
 215 220 225
 His Lys Arg Arg Pro Lys Val Lys Arg Ser Lys Lys Arg Arg Lys
 230 235 240
 Glu Ala Ser Ser Ser Glu Glu Pro Arg Asn Lys His Ala Met Asn
 245 250 255
 Pro Lys Gly His Ser Glu Arg Ser Asp Thr Asn Glu Lys Arg Ser
 260 265 270
 Val Asp Ser Ser Ala Lys Arg Glu Lys Pro Val Val Arg Pro Glu
 275 280 285
 Glu Ile Pro Pro Val Pro Glu Asn Arg Phe Leu Leu Arg Arg Asp
 290 295 300
 Met Pro Val Val Thr Ala Glu Pro Glu Pro Ile Pro Asp Val Ala
 305 310 315
 Pro Ile Val Ser Asp Gln Lys Pro Ser Val Ser Lys Ser Gly Arg
 320 325 330
 Lys Ile Lys Gly Arg Gly Thr Ile Arg Tyr His Thr Pro Pro Arg
 335 340 345
 Ser Arg Ser Cys Ser Glu Ser Asp Asp Asp Ser Ser Glu Thr
 350 355 360
 Pro Pro His Trp Lys Glu Glu Met Gln Arg Leu Arg Ala Tyr Arg
 365 370 375
 Pro Pro Ser Gly Glu Lys Trp Ser Lys Gly Asp Lys Leu Ser Asp
 380 385 390
 Pro Cys Ser Ser Arg Trp Asp Glu Arg Ser Leu Ser Gln Arg Ser
 395 400 405
 Arg Ser Trp Ser Tyr Asn Gly Tyr Tyr Ser Asp Leu Ser Thr Ala
 410 415 420
 Arg His Ser Gly His His Lys Lys Arg Arg Lys Glu Lys Lys Val
 425 430 435
 Lys His Lys Lys Lys Gly Lys Lys Gln Lys His Cys Arg Arg His
 440 445 450
 Lys Gln Thr Lys Lys Arg Arg Ile Leu Ile Pro Ser Asp Ile Glu
 455 460 465
 Ser Ser Lys Ser Ser Thr Arg Arg Met Lys Ser Ser Cys Asp Arg
 470 475 480
 Glu Arg Ser Ser Arg Ser Ser Ser Leu Ser Ser His His Ser Ser
 485 490 495
 Lys Arg Asp Trp Ser Lys Ser Asp Lys Asp Val Gln Ser Ser Leu
 500 505 510
 Thr His Ser Ser Arg Asp Ser Tyr Arg Ser Lys Ser His Ser Gln
 515 520 525
 Ser Tyr Ser Arg Gly Ser Ser Arg Ser Arg Thr Ala Ser Lys Ser
 530 535 540
 Ser Ser His Ser Arg Ser Arg Ser Lys Ser Arg Ser Ser Ser Lys
 545 550 555
 Ser Gly His Arg Lys Arg Ala Ser Lys Ser Pro Arg Lys Thr Ala
 560 565 570
 Ser Gln Leu Ser Glu Asn Lys Pro Val Lys Thr Glu Pro Leu Arg
 575 580 585

WO 02/097032

PCT/US02/11152

Ala Thr Met Ala Gln Asn Glu Asn Val Val Val Gln Pro Val Val
 590 595 600
 Ala Glu Asn Ile Pro Val Ile Pro Leu Ser Asp Ser Pro Pro Pro
 605 610 615
 Ser Arg Trp Lys Pro Gly Gln Lys Pro Trp Lys Pro Ser Tyr Glu
 620 625 630
 Arg Ile Gln Glu Met Lys Ala Lys Thr Thr His Leu Leu Pro Ile
 635 640 645
 Gln Ser Thr Tyr Ser Leu Ala Asn Ile Lys Glu Thr Gly Ser Ser
 650 655 660
 Ser Ser Tyr His Lys Arg Glu Lys Asn Ser Glu Ser Asp Gln Ser
 665 670 675
 Thr Tyr Ser Lys Tyr Ser Asp Arg Ser Ser Glu Ser Ser Pro Arg
 680 685 690
 Ser Arg Ser Arg Ser Ser Arg Ser Arg Ser Tyr Ser Arg Ser Tyr
 695 700 705
 Thr Arg Ser Arg Ser Leu Ala Ser Ser His Ser Arg Ser Arg Ser
 710 715 720
 Pro Ser Ser Arg Ser His Ser Arg Asn Lys Tyr Ser Asp His Ser
 725 730 735
 Gln Cys Ser Arg Ser Ser Ser Tyr Thr Ser Ile Ser Ser Asp Asp
 740 745 750
 Gly Arg Arg Ala Lys Arg Arg Leu Arg Ser Ser Gly Lys Lys Asn
 755 760 765
 Ser Val Ser His Lys Lys His Ser Ser Ser Glu Lys Thr Leu
 770 775 780
 His Ser Lys Tyr Val Lys Gly Arg Asp Ser Ser Cys Val Arg
 785 790 795
 Lys Tyr Ser Glu Ser Arg Ser Ser Leu Asp Tyr Ser Ser Asp Ser
 800 805 810
 Glu Gln Ser Ser Val Gln Ala Thr Gln Ser Ala Gln Glu Lys Glu
 815 820 825
 Lys Gln Gly Gln Met Glu Arg Thr His Asn Lys Gln Glu Lys Asn
 830 835 840
 Arg Gly Glu Glu Lys Ser Lys Ser Glu Arg Glu Cys Pro His Ser
 845 850 855
 Lys Lys Arg Thr Leu Lys Glu Asn Leu Ser Asp His Leu Arg Asn
 860 865 870
 Gly Ser Lys Pro Lys Arg Lys Asn Tyr Ala Gly Ser Lys Trp Asp
 875 880 885
 Ser Glu Ser Asn Ser Glu Arg Asp Val Thr Lys Asn Ser Lys Asn
 890 895 900
 Asp Ser His Pro Ser Ser Asp Lys Glu Glu Gly Glu Ala Thr Ser
 905 910 915
 Asp Ser Glu Ser Glu Val Ser Glu Ile His Ile Lys Val Lys Pro
 920 925 930
 Thr Thr Lys Ser Ser Thr Asn Thr Ser Leu Pro Asp Asp Asn Gly
 935 940 945
 Ala Trp Lys Ser Ser Lys Gln Arg Thr Ser Thr Ser Asp Ser Glu
 950 955 960
 Gly Ser Cys Ser Asn Ser Glu Asn Asn Arg Gly Lys Pro Gln Lys
 965 970 975
 His Lys His Gly Ser Lys Glu Asn Leu Lys Arg Glu His Thr Lys
 980 985 990
 Lys Val Lys Glu Lys Leu Lys Gly Lys Lys Asp Lys Lys His Lys
 995 1000 1005

WO 02/097032

PCT/US02/11152

Ala Pro Lys Arg Lys Gln Ala Phe His Trp Gln Pro Pro Leu Glu
 1010 1015 1020
 Phe Gly Glu Glu Glu Glu Glu Ile Asp Asp Lys Gln Val Thr
 1025 1030 1035
 Gln Glu Ser Lys Glu Lys Lys Val Ser Glu Asn Asn Glu Thr Ile
 1040 1045 1050
 Lys Asp Asn Ile Leu Lys Thr Glu Lys Ser Ser Glu Glu Asp Leu
 1055 1060 1065
 Ser Gly Lys His Asp Thr Val Thr Val Ser Ser Asp Leu Asp Gln
 1070 1075 1080
 Phe Thr Lys Asp Asp Ser Lys Leu Ser Ile Ser Pro Thr Ala Leu
 1085 1090 1095
 Asn Thr Glu Glu Asn Val Ala Cys Leu Gln Asn Ile Gln His Val
 1100 1105 1110
 Glu Glu Ser Val Pro Asn Gly Val Glu Asp Val Leu Gln Thr Asp
 1115 1120 1125
 Asp Asn Met Glu Ile Cys Thr Pro Asp Arg Ser Ser Pro Ala Lys
 1130 1135 1140
 Val Glu Glu Thr Ser Pro Leu Gly Asn Ala Arg Leu Asp Thr Pro
 1145 1150 1155
 Asp Ile Asn Ile Val Leu Lys Gln Asp Met Ala Thr Glu His Pro
 1160 1165 1170
 Gln Ala Glu Val Val Lys Gln Glu Ser Ser Met Ser Glu Ser Lys
 1175 1180 1185
 Val Leu Gly Glu Val Gly Lys Gln Asp Ser Ser Ser Ala Ser Leu
 1190 1195 1200
 Ala Ser Ala Gly Glu Ser Thr Gly Lys Lys Glu Val Ala Glu Lys
 1205 1210 1215
 Ser Gln Ile Asn Leu Ile Asp Lys Lys Trp Lys Pro Leu Gln Gly
 1220 1225 1230
 Val Gly Asn Leu Ala Ala Pro Asn Ala Ala Thr Ser Ser Ala Val
 1235 1240 1245
 Glu Val Lys Val Leu Thr Thr Val Pro Glu Met Lys Pro Gln Gly
 1250 1255 1260
 Leu Arg Ile Glu Ile Lys Ser Lys Asn Lys Val Arg Pro Gly Ser
 1265 1270 1275
 Leu Phe Asp Glu Val Arg Lys Thr Ala Arg Leu Asn Arg Arg Pro
 1280 1285 1290
 Arg Asn Gln Glu Ser Ser Ser Asp Glu Gln Thr Pro Ser Arg Asp
 1295 1300 1305
 Asp Asp Ser Gln Ser Arg Ser Pro Ser Arg Ser Arg Ser Lys Ser
 1310 1315 1320
 Glu Thr Lys Ser Arg His Arg Thr Arg Ser Val Ser Tyr Ser His
 1325 1330 1335
 Ser Arg Ser Arg Ser Arg Ser Ser Thr Ser Ser Tyr Arg Ser Arg
 1340 1345 1350
 Ser Tyr Ser Arg Ser Arg Ser Arg Gly Trp Tyr Ser Arg Gly Arg
 1355 1360 1365
 Thr Arg Ser Arg Ser Ser Tyr Arg Ser Tyr Lys Ser His Arg
 1370 1375 1380
 Thr Ser Ser Arg Ser Arg Ser Ser Ser Tyr Asp Pro His
 1385 1390 1395
 Ser Arg Ser Ser Arg Ser Tyr Thr Tyr Asp Ser Tyr Tyr Ser Arg
 1400 1405 1410
 Ser Arg Ser Arg Ser Arg Ser Gln Arg Ser Asp Ser Tyr His Arg
 1415 1420 1425

WO 02/097032

PCT/US02/11152

Gly Arg Ser Tyr Asn Arg Arg Ser Arg Ser Cys Arg Ser Tyr Gly
1430 1435 1440
Ser Asp Ser Glu Ser Asp Arg Ser Tyr Ser His His Arg Ser Pro
1445 1450 1455
Ser Glu Ser Ser Arg Tyr Ser
1460

<210> 18
<211> 329
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7495168CD1

<400> 18
Met Ser Ala Glu Ala Ser Gly Pro Ala Ala Ala Ala Pro Ser
1 5 10 15
Leu Glu Ala Pro Lys Pro Ser Gly Leu Glu Pro Gly Pro Ala Ala
20 25 30
Tyr Gly Leu Lys Pro Leu Thr Pro Asn Ser Lys Tyr Val Lys Leu
35 40 45
Asn Val Gly Gly Ser Leu His Tyr Thr Thr Leu Arg Thr Leu Thr
50 55 60
Gly Gln Asp Thr Met Leu Lys Ala Met Phe Ser Gly Arg Val Glu
65 70 75
Val Leu Thr Asp Ala Gly Gly Trp Val Leu Ile Asp Arg Ser Gly
80 85 90
Arg His Phe Gly Thr Ile Leu Asn Tyr Leu Arg Asp Gly Ser Val
95 100 105
Pro Leu Pro Glu Ser Thr Arg Glu Leu Gly Glu Leu Leu Gly Glu
110 115 120
Ala Arg Tyr Tyr Leu Val Gln Gly Leu Ile Glu Asp Cys Gln Leu
125 130 135
Ala Leu Gln Gln Lys Arg Glu Thr Leu Ser Pro Leu Cys Leu Ile
140 145 150
Pro Met Val Thr Ser Pro Arg Glu Glu Gln Gln Leu Leu Ala Ser
155 160 165
Thr Ser Lys Pro Val Val Lys Leu Leu His Asn Arg Ser Asn Asn
170 175 180
Lys Tyr Ser Tyr Thr Ser Thr Ser Asp Asp Asn Leu Leu Lys Asn
185 190 195
Ile Glu Leu Phe Asp Lys Leu Ala Leu Arg Phe His Gly Arg Leu
200 205 210
Leu Phe Leu Lys Asp Val Leu Gly Asp Glu Ile Cys Cys Trp Ser
215 220 225
Phe Tyr Gly Gln Gly Arg Lys Ile Ala Glu Val Cys Cys Thr Ser
230 235 240
Ile Val Tyr Ala Thr Glu Lys Lys Gln Thr Lys Val Glu Phe Pro
245 250 255
Glu Ala Arg Ile Phe Glu Glu Thr Leu Asn Ile Leu Ile Tyr Glu
260 265 270
Thr Pro Arg Gly Pro Asp Pro Ala Leu Leu Glu Ala Thr Gly Gly
275 280 285
Ala Ala Gly Ala Gly Gly Ala Gly Arg Gly Glu Asp Glu Glu Asn

WO 02/097032

PCT/US02/11152

Arg	Glu	His	Arg	Val	Arg	Arg	Ile	His	Val	Arg	Arg	His	Ile	Thr
305									310					315
His	Asp	Glu	Arg	Pro	His	Gly	Gln	Gln	Ile	Val	Phe	Lys	Asp	
320									325					

<210> 19

<211> 476

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7483131CD1

<400> 19

Met	Val	Lys	Leu	Ile	His	Thr	Leu	Ala	Asp	His	Gly	Asp	Asp	Val
1			5						10					15
Asn	Cys	Cys	Ala	Phe	Ser	Phe	Ser	Leu	Leu	Ala	Thr	Cys	Ser	Leu
			20						25					30
Asp	Lys	Thr	Ile	Arg	Leu	Tyr	Ser	Leu	Arg	Asp	Phe	Thr	Glu	Leu
			35						40					45
Pro	His	Ser	Pro	Leu	Lys	Phe	His	Thr	Tyr	Ala	Val	His	Cys	Cys
			50						55					60
Cys	Phe	Ser	Pro	Ser	Gly	His	Ile	Leu	Ala	Ser	Cys	Ser	Thr	Asp
			65						70					75
Gly	Thr	Thr	Val	Leu	Trp	Asn	Thr	Glu	Asn	Gly	Gln	Met	Leu	Ala
			80						85					90
Val	Met	Glu	Gln	Pro	Ser	Gly	Ser	Pro	Val	Arg	Val	Cys	Gln	Phe
			95						100					105
Ser	Pro	Asp	Ser	Thr	Cys	Leu	Ala	Ser	Gly	Ala	Ala	Asp	Gly	Thr
			110						115					120
Val	Val	Leu	Trp	Asn	Ala	Gln	Ser	Tyr	Lys	Leu	Tyr	Arg	Cys	Gly
			125						130					135
Ser	Val	Lys	Asp	Gly	Ser	Leu	Ala	Ala	Cys	Ala	Phe	Ser	Pro	Asn
			140						145					150
Gly	Ser	Phe	Phe	Val	Thr	Gly	Ser	Ser	Cys	Gly	Asp	Leu	Thr	Val
			155						160					165
Trp	Asp	Asp	Lys	Met	Arg	Cys	Leu	His	Ser	Glu	Lys	Ala	His	Asp
			170						175					180
Leu	Gly	Ile	Thr	Cys	Cys	Asp	Phe	Ser	Ser	Gln	Pro	Val	Ser	Asp
			185						190					195
Gly	Glu	Gln	Gly	Leu	Gln	Phe	Phe	Arg	Leu	Ala	Ser	Cys	Gly	Gln
			200						205					210
Asp	Cys	Gln	Val	Lys	Ile	Trp	Ile	Val	Ser	Phe	Thr	His	Ile	Leu
			215						220					225
Gly	Phe	Glu	Leu	Lys	Tyr	Lys	Ser	Thr	Leu	Ser	Gly	His	Cys	Ala
			230						235					240
Pro	Val	Leu	Ala	Cys	Ala	Phe	Ser	His	Asp	Gly	Gln	Met	Leu	Val
			245						250					255
Ser	Gly	Ser	Val	Asp	Lys	Ser	Val	Ile	Val	Tyr	Asp	Thr	Asn	Thr
			260						265					270
Glu	Asn	Ile	Leu	His	Thr	Leu	Thr	Gln	His	Thr	Arg	Tyr	Val	Thr
			275						280					285
Thr	Cys	Ala	Phe	Ala	Pro	Asn	Thr	Leu	Leu	Leu	Ala	Thr	Gly	Ser
			290						295					300

WO 02/097032

PCT/US02/11152

```

Met Asp Lys Thr Val Asn Ile Trp Gln Phe Asp Leu Glu Thr Leu
305 310 315
Cys Gln Ala Arg Ser Thr Glu His Gln Leu Lys Gln Phe Thr Glu
320 325 330
Asp Trp Ser Glu Glu Asp Val Ser Thr Trp Leu Cys Ala Gln Asp
335 340 345
Leu Lys Asp Leu Val Gly Ile Phe Lys Met Asn Asn Ile Asp Gly
350 355 360
Lys Glu Leu Leu Asn Leu Thr Lys Glu Ser Leu Ala Asp Asp Leu
365 370 375
Lys Ile Glu Ser Leu Gly Leu Arg Ser Lys Val Leu Arg Lys Ile
380 385 390
Glu Glu Leu Arg Thr Lys Val Lys Ser Leu Ser Ser Gly Ile Pro
395 400 405
Asp Glu Phe Ile Cys Pro Ile Thr Arg Glu Leu Met Lys Asp Pro
410 415 420
Val Ile Ala Ser Asp Gly Tyr Ser Tyr Glu Lys Glu Ala Met Glu
425 430 435
Asn Trp Ile Ser Lys Lys Lys Arg Thr Ser Pro Met Thr Asn Leu
440 445 450
Val Leu Pro Ser Ala Val Leu Thr Pro Asn Arg Thr Leu Lys Met
455 460 465
Ala Ile Asn Arg Trp Leu Glu Thr His Gln Lys
470 475

```

<210> 20

<211> 485

<212> PPT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 4558650CD1

<400> 20

```

Met Ala Leu Pro Pro Phe Phe Gly Gln Gly Arg Pro Gly Pro Pro
1 5 10 15
Pro Pro Gln Pro Pro Pro Pro Ala Pro Phe Gly Cys Pro Pro Pro
20 25 30
Pro Leu Pro Ser Pro Ala Phe Pro Pro Pro Leu Pro Gln Arg Pro
35 40 45
Gly Pro Phe Pro Gly Ala Ser Ala Pro Phe Leu Gln Pro Pro Leu
50 55 60
Ala Leu Gln Pro Arg Ala Ser Ala Glu Ala Ser Arg Gly Gly Gly
65 70 75
Gly Ala Gly Ala Phe Tyr Pro Val Pro Pro Pro Pro Leu Pro Pro
80 85 90
Pro Pro Pro Gln Cys Arg Pro Phe Pro Gly Thr Asp Ala Gly Glu
95 100 105
Arg Pro Arg Pro Pro Pro Pro Gly Pro Gly Pro Pro Trp Ser Pro
110 115 120
Arg Trp Pro Glu Ala Pro Pro Pro Pro Ala Asp Val Leu Gly Asp
125 130 135
Ala Ala Leu Gln Arg Leu Arg Asp Arg Gln Trp Leu Glu Ala Val
140 145 150
Phe Gly Thr Pro Arg Arg Ala Gly Cys Pro Val Pro Gln Arg Thr

```

WO 02/097032

PCT/US02/11152

155 160 165
His Ala Gly Pro Ser Leu Gly Glu Val Arg Ala Arg Leu Leu Arg
170 175 180
Ala Leu Arg Leu Val Arg Arg Leu Arg Gly Leu Ser Gln Ala Leu
185 190 195
Arg Glu Ala Glu Ala Asp Gly Ala Ala Trp Val Leu Leu Tyr Ser
200 205 210
Gln Thr Ala Pro Leu Arg Ala Glu Leu Ala Glu Arg Leu Gln Pro
215 220 225
Leu Thr Gln Ala Ala Tyr Val Gly Glu Ala Arg Arg Arg Leu Glu
230 235 240
Arg Val Arg Arg Arg Arg Leu Arg Leu Arg Glu Arg Ala Arg Glu
245 250 255
Arg Glu Ala Glu Arg Glu Ala Glu Ala Ala Arg Ala Val Glu Arg
260 265 270
Glu Gln Glu Ile Asp Arg Trp Arg Val Lys Cys Val Gln Glu Val
275 280 285
Glu Glu Lys Lys Arg Glu Gln Glu Leu Lys Ala Ala Ala Asp Gly
290 295 300
Val Leu Ser Glu Val Arg Lys Lys Gln Ala Asp Thr Lys Arg Met
305 310 315
Val Asp Ile Leu Arg Ala Leu Glu Lys Leu Arg Lys Leu Arg Lys
320 325 330
Glu Ala Ala Ala Arg Lys Gly Val Cys Pro Pro Ala Ser Ala Asp
335 340 345
Glu Thr Phe Thr His His Leu Gln Arg Leu Arg Lys Leu Ile Lys
350 355 360
Lys Arg Ser Glu Leu Tyr Glu Ala Glu Glu Arg Ala Leu Arg Val
365 370 375
Met Leu Glu Gly Glu Gln Glu Glu Glu Arg Lys Arg Glu Leu Glu
380 385 390
Lys Lys Gln Arg Lys Glu Glu Glu Lys Ile Leu Leu Gln Lys Arg
395 400 405
Glu Ile Glu Ser Lys Leu Phe Gly Asp Pro Asp Glu Phe Pro Leu
410 415 420
Ala His Leu Leu Glu Pro Phe Arg Gln Tyr Tyr Leu Gln Ala Glu
425 430 435
His Ser Leu Pro Ala Leu Ile Gln Ile Arg His Asp Trp Asp Gln
440 445 450
Tyr Leu Val Pro Ser Asp His Pro Lys Gly Asn Phe Val Pro Gln
455 460 465
Gly Trp Val Leu Pro Pro Leu Pro Ser Asn Asp Ile Trp Ala Thr
470 475 480
Ala Val Lys Leu His
485

<210> 21

<211> 406

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7506195CD1

<400> 21

WO 02/097032

PCT/US02/11152

```

Met Ala Leu Pro Pro Phe Phe Gly Gln Gly Arg Pro Gly Pro Pro
1      5      10      15
Pro Pro Gln Pro Pro Pro Ala Pro Phe Gly Cys Pro Pro Pro
20      25      30
Pro Leu Pro Ser Pro Ala Phe Pro Pro Pro Leu Pro Gln Arg Pro
35      40      45
Gly Pro Phe Pro Gly Ala Ser Ala Pro Phe Leu Gln Pro Pro Leu
50      55      60
Ala Leu Gln Pro Arg Ala Ser Ala Glu Ala Ser Arg Gly Gly Gly
65      70      75
Gly Ala Gly Ala Phe Tyr Pro Val Pro Pro Pro Leu Pro Pro
80      85      90
Pro Pro Pro Gln Cys Arg Pro Phe Pro Gly Thr Asp Ala Gly Glu
95      100      105
Arg Pro Arg Pro Pro Pro Gly Pro Gly Pro Pro Trp Ser Pro
110      115      120
Arg Trp Pro Glu Ala Pro Pro Pro Ala Asp Val Leu Gly Asp
125      130      135
Ala Ala Leu Gln Arg Leu Arg Asp Arg Gln Trp Leu Glu Ala Val
140      145      150
Phe Gly Thr Pro Arg Arg Ala Gly Cys Pro Val Pro Gln Arg Thr
155      160      165
His Ala Gly Pro Ser Leu Gly Glu Val Arg Ala Arg Leu Leu Arg
170      175      180
Ala Leu Arg Leu Val Arg Arg Leu Arg Gly Leu Ser Gln Ala Leu
185      190      195
Arg Glu Ala Glu Ala Asp Gly Ala Ala Trp Val Leu Leu Tyr Ser
200      205      210
Gln Thr Ala Pro Leu Arg Ala Glu Leu Ala Glu Arg Leu Gln Pro
215      220      225
Leu Thr Gln Ala Ala Tyr Val Gly Glu Ala Arg Arg Arg Leu Glu
230      235      240
Arg Val Arg Arg Arg Arg Leu Arg Leu Arg Glu Arg Ala Arg Glu
245      250      255
Arg Glu Ala Glu Arg Glu Ala Glu Ala Ala Arg Ala Val Glu Arg
260      265      270
Glu Gln Glu Ile Asp Arg Trp Arg Val Lys Cys Val Gln Glu Val
275      280      285
Glu Glu Lys Lys Arg Glu Gln Glu Leu Lys Ala Ala Ala Asp Gly
290      295      300
Val Leu Ser Glu Val Arg Lys Lys Gln Ala Asp Thr Lys Arg Met
305      310      315
Val Asp Ile Leu Arg Ala Leu Glu Lys Leu Arg Lys Leu Arg Lys
320      325      330
Glu Ala Ala Ala Arg Lys Asp Glu Phe Pro Leu Ala His Leu Leu
335      340      345
Glu Pro Phe Arg Gln Tyr Tyr Leu Gln Ala Glu His Ser Leu Pro
350      355      360
Ala Leu Ile Gln Ile Arg His Asp Trp Asp Gln Tyr Leu Val Pro
365      370      375
Ser Asp His Pro Lys Gly Asn Phe Val Pro Gln Gly Trp Val Leu
380      385      390
Pro Pro Leu Pro Ser Asn Asp Ile Trp Ala Thr Ala Val Lys Leu
395      400      405
His

```

WO 02/097032

PCT/US02/11152

<210> 22

<211> 7742

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1351608CB1

<400> 22

```

ggctactgaga ggaaggaaat cagcgccggg ttaccccaga cccctcagcg tctggcactc 60
tgggtgggac agcaagttga tttttatact gotttcttgc atcatttggc acaattggg 120
ccagaaattt actttgttga aatggaccca gacttggaaa agcaggagga aagtgtacaa 180
atgtcaaatat tcactccact ggaatcgtac ttatttggag aagatccaga tatttgccta 240
gagaaatttga agcacagcgg agcatttccg ctttggggga gggttttcaa aagtggagag 300
acaacctatt cttgcaggga ttgtgcaatt gatccaacat ggtactctgt tatggactgc 360
ttccaggaca gtgttcataa aaatcactgt tacaagatgc atactcttac tggaggaggg 420
tctgtgact gtggagacac agaggcatgg aaaaactggc ctttttgtgt aaatcatgaa 480
cctggaaagc caggtactat aaaaagagaat tccgctgtgc cgttgaatga agaggttaatt 540
ttccaagcca ggaataatt tccctcagtg ataaaatag atagataatg gactataatg 600
gaagaggaaa aagaactgcc tccctgaactc cagataaggg agaaaaatga aagatactat 660
tgtgtccttt tcaatgatga acacattcca tatgaccacg tcatatacag cctacaaaga 720
gctcttgact gtgagctcgc agaggcccg ttgcatacca ctgocattga caaagaggg 780
cgtcggcggt taaagccggg agcttatgct gcttgcagg aagcaagga agatataaag 840
agtcaattccg aaaaatgtct tcaacatcca ctctatgtag aagtattaca ctcagagatt 900
atggctcctc agaaatttgc tttgcgtctt ggttccctga tgaacaaaat tatgagctat 960
tcaagtgtact ttaggcagat cttttgccaa gcatgcctta gagaagaacc tgaactcgag 1020
aatccctctc tcaaaagcag gttaatgctt tgggatgc aa agctttataa aggtgccctc 1080
aagtcccttc atgaattgat ctccagcagt ttttttatgg agatggaata caaaaaactc 1140
tttgtctatg aatttgtgaa gtattataaa caactgcaga aagaatatat cagtgtgat 1200
catgacagaa gtatctctat aactgcactt tcaattcaga tgtttactgt tctactctgt 1260
gtctcagcac ttattgaaga gcgaatgtt atctctgcca ttaactgaac tctgtataga 1320
gttttaactg agtacttggg caggaacaaat aaattcaact tccagggtta tagcaggac 1380
aaattgggaa gagtatatgc agtaatatgt gaactaaagt atactctgat cagcaaaccc 1440
acaatatgga cagaagattt aagaatpccg ttccctgaag gttttcgac ttttttgag 1500
attcttaact gtatgcaggg aatggaagaa atccgaagac aggttgggca acacattgaa 1560
gtggatccgt attgggaggg tgcatttgc atacagatgc aattgaagaa tattttactc 1620
atgttcccaag agtgggtgtg ttgtgatgaa gaactcttac ttgtggctta taagaatgt 1680
cacaaagctg tgatgaggtg cagtaccagt ttcatatcta gtgacaagac agtagtaca 1740
tcgtgtggac atagtttgga aacaaagtc tccagagtat ctgaggatct tgaagcata 1800
catctgccac tctctaggac ccttgcctgt ctcatgtac gtttaagcag gctgggtgct 1860
gtttcaagac tgcattgaatt tgtgtctttt gaggacttcc aagttagagg actagtggaa 1920
tatcctttac gttgtctggt gttggttgc caggttgttg ctgagatgtg cgaagaaat 1980
ggactgtctc ttattagcca ggtgttttat taocaaagat ttaagtgcag agaagaatg 2040
tatgataaag atatcatcat gcttcagatt ggtgcactct taatggatcc caataagttc 2100
ttgttactgg tacttcagag gtatgaactt gccagagctt ttaacaagac catatctaca 2160
aaagccaggg atttgatga acaatataat acactaatag aagaatgct tcaagttctc 2220
atctatatgg tgggtgagcg ttatgtacct ggagtgggaa atgtgaccaa agaagaggtc 2280
acaatgagag aaatcattca cttgttttgc attgaaccca tgcacacag tgcatttgc 2340
aaaaatttcc ctgagaatga aaataatgaa actggcttag agaattgcat aaacaagtg 2400
gccacattta agaaaccagg tctatccagg cactggagttt atgaactaaa agatgaatca 2460
ctgaagagat tcaatatgta cttttatcat taactcaaaa cccagcatag caagcctgaa 2520
catatgcaga agaaagagag aaaaacaaga aacaagatg aagcatttgc gccaccca 2580
ctccctgaat tctgcccctg tttcagcaaa gtgattaacc ttctcaactg tgatatcatg 2640
atgtcatttc tcaagaccgt atttgagcgg gcaatagaca cagattctaa cttgtggacc 2700
gaagggtatc tccaaatggc ttttcatatt ctggcatttg gtttaactaga agagaagcaa 2760

```


WO 02/097032

PCT/US02/11152

cagcttcaaa aagctctga agaagaagta acatttgact tttatcataa ggottcaaga 2820
 ttgggaagtt cagccatgaa tatacaaatg cttttggaaa aactcaaggg aattccccag 2880
 ttaagaagccc agaaggacat gataacgtgg atacttcaga tgtttgacac agtgaagcga 2940
 ttaagagaaaa aatcttgttt aattgtagca accacacag gatcggaatc tattaagaat 3000
 gatgagatta ctcatgataa agaaaaagca gaacgaaaaa gaaaagctga agctgctagg 3060
 ctacatctcc agaagatcat ggctcagatg tctgctttac agaaaaactt cattgaact 3120
 cataaactca tgtatgacaa tacatcagaa atgctctggg aagaagattc cattatggag 3180
 gaagagagca cccagcagat cagtgaactac tctagaattg ctttgggtcc taacgggggt 3240
 ccatctgtta ctgaaaagga ggtgctgaag tgcactcttt gccaaagaga acagagggtg 3300
 aaaaatagaaa ataattgccat ggtattatcg gctgtgtctc agaatctac tgccttaac 3360
 cagcaccggg gaaaacccat agaacctctc ggagaagccc tagaccactc ttctatggat 3420
 ccagacttgg catatggaa cttatcagga agctgtggto atgtaatgca ccgagtggtc 3480
 tggcagaagt attttgaagc tgtacagctg agctctcagc agcgattcca tgttgacct 3540
 tttgacttgg aaagtggaga atatctttgc cctctttgca aatctctgtg caactctgtg 3600
 atccccatta ttcctttgca acctcaaaag ataacacgtg agaagcaga tgcctctgtc 3660
 caacttttga ccttggcagc gtggatacag actgttcttg ccagaatata aggtataat 3720
 atagacatg ctaaggagga aaacccaatt cctattttct ttaatacagg aatgggagt 3780
 tctacttttg agtccattc catctgagt tttggcgttg agtcttgat taaatattca 3840
 aatgcacatc aggaatgggt tattctcttt gccacaaca tttatagaat tggattgaa 3900
 gtgccacctg atgaaaagga tctctgagtc cccatgctga cctggagcac ctgcgcttcc 3960
 actatccagg caattgaaaa tctattggga gatgaaggaa aacctctgtt tggagcact 4020
 caaaataggc agcataatgg tctgaaagca ttaatgcagt ttcagttgc acagaggatt 4080
 accgtctctc aggtctgat acagaaacat ctggtctgtc tctatcagt tgtttctctc 4140
 aacataaaat cagaagatc accatgctct ctgtctatag atctgtttca tgttttgggt 4200
 ggtgctgtgt tagcattccc atcctgttat tgggatgacc ctgttgatct gcagcctctc 4260
 tctgttagtt cttctataaa caacctttat cttctccatt tcatcaccat ggcaacactg 4320
 ctccagatac taactacagt agacacagcg ctaccccttg ctcaggttca agaagacagt 4380
 gaagaggctc attcgcgctc tctctttctt gcagaaatt ctaataatac aagtggtctc 4440
 attgggtgtg atattctctg ctggtatttg tgggtctcac tgaagaatgg catcacctc 4500
 tatctctctg gtgctgact gtttttccac tatttaactg ggttaactcc gctcaggaa 4560
 ctgcatacca attctgcaga aggaagatc agtgcaactc gtactatct atcttactc 4620
 acaaatctgt tctgtctctt ccaggaaatc tgggatactg taaggccctt gctccagggt 4680
 tgggtgtcag atctgtctct actaaactgt ttgaagcaaa aaacacactg ggtcaggtac 4740
 cctagaaaaa gaaatagttt gatagagctt cctgatgact atagctgctc cctgaatcaa 4800
 gttctctact taagggtccc aoggtctgca gatgatgagc gaaagcattc tgtctctctc 4860
 cttttctgtg gggctatact atgtctctag aacatttgct gccaggaaat tgtgaacggg 4920
 gaagagggtg gagcttgcac ttttccagca ctctactgtg gagcggaggt ctgcattttc 4980
 ctaaaaatca gagaatgccc agtgggtctg gttgaaggtc aagccagagg ctgtgctcat 5040
 ccagctctct acttggatga atatgggaaa acagacccct gctcgaagag gggcaacccc 5100
 cttcatttat ctgtgagcg gtatcgggaag ctccatttgg tctggcaaca acactgcatt 5160
 atagaagaga ttgctaggag ccaagagact aatcagatgt tatttggatt caactggcag 5220
 ttactgtgag ctccactct gctcagagc aatcaccaat gacgacagta gtaaggctg 5280
 attcaaaatt atgaaaaact tctgagggc tgggaaagta ttggagggtc ttttgcctca 5340
 tgtccagggt cacttacatc aataaatat tcttcaatgg agtattgctc tcaattagca 5400
 aacatattgt tcacaggaaa aaaggacata gatcaatctg ttttatgtgc tagtatttcc 5460
 aggaatttat tcccttcat aatttgtctc atttcatttt atttcaatca ctggtagat 5520
 gaagtcaagt caaacagttg tagacatttt atgtgttgggt taactcttct gccattttgt 5580
 atttgggtgt ttccccca gtttagttca actgacattg gatcactgac aaaattctaa 5640
 taactctgta tagtcttctc tgcagttaaa gaagaattgc agaaacatg caataactc 5700
 gggaaagatt ccaaaaataa attttttatt atttctcttt taaggaaata cccataatg 5760
 gccacctctg gctatccaca caaattaaac tcaatctcta tgtggacaga ggtatgttc 5820
 tgcacatctg gaaaagcttt ttctcactg taggctcaca gaaaagttag ggtaatttat 5880
 ttgttatbca ttctgacggc tacaagagc ttgcagttct cactctgac taccagtagc 5940
 ttgttgaggt ttgaaataa taactgacat ttccaaagg caaatctcat tctgcaagg 6000
 gatctggc caactctgtt tgactctcag aaacctctg taattctgat gtaaaaactg 6060
 tagaatgaag atgagaaaat tctgcgaatg agtggatcat gacaactgta aattagaaca 6120

WO 02/097032

PCT/US02/11152

```
atcagattta aaccaattcc gcaatcttct atctctttgt aaaagacaaa tctttgatgt 6180
tgtctgtgtg caaccttttc ataaactctg gttttatgac tagtacaaa caccacaaaa 6240
gccatgtgat caatagctctg tctcctgtta taacatgctg tggttgagcc atcttgttta 6300
taaatcaatag agctctctctg aatttctgca tagacttctt ggttctctgg ttttcttttt 6360
tgtatcaaga gattgtgata taaaacagca gaagataaat ggaacacctc cattttaact 6420
tacgttgttt ctgggttaat gttagaacct tgaagatgc attcaagac tgtattttat 6480
tttgcctctg gctattagtg tctcacatat gtgtgttaat gttttcttac cttctttttg 6540
ctacgcaaaag gcaagcaagt aaaaatatatt tgcctagtgga ttagtgtatgc acatttgggg 6600
ctagatctttt ttggtacttt tatgtaaaag aaagtgtggt ttgcagtaag ggaattggcat 6660
gagcaggcgt cagaatccca atcctgattt tctacttgaa taattacaat tcagaagga 6720
tctggtaaaa tagatacatg tctagtgaac aatttgaac aataacaggt aaggatcagg 6780
aaattcagta ttcagtttgt cagattttgc agaatgatga aagtatttga acagtgtgtg 6840
ttgtctctta tataattgta ttgagtgatg tgtttgactg ggaactctgg gctagaatag 6900
gaacagaag atactgactt ctacccctaat agatgggccc caatttagca aagataaact 6960
gactttattt ttactctctt ttatattaac ttaacaaatt ctggagttag gctctcaaga 7020
ggacagaggg actgtctctg aatggccagc cagaccttta ctgccaaaga acccatttca 7080
tattgcgttc cactgattga gattgattca gatttttgca ctgtagatga gcttagttct 7140
cagtgctgcc ccaagcccca gggatttctc attatgttca aatgtctag tgatttacct 7200
taatcattgc aaacaattat gcttatgaag tttaacttca acaagcaac tgagtacct 7260
tattttcttt agttagtat gtgaaggcac tgyttcaaca ggaatggctcc agaatctgtg 7320
ttttctaatg ttgtgtaagg ggtagttag aattttaatg atatggtgaa gaaaaatata 7380
tctgtataat taatttatta tttgtgtgta tgggctgtga gtctaccttt tagtggcat 7440
ttgtcatctc ataacaacta tgcatttttg ttcactgtga tgaatgacta tatttagtga 7500
gtgcacatg tttataccac tgattcaaat tccatccatg atgaagttaa acaataaatg 7560
catatattga taacttttat tgcacaaatg taaatttaaa acttgtataa tgtttctgtg 7620
ctttttaaaa taaaatatat gtgtatatct aaaaagaaaa aaaaaaaaaa aaaaaaaa 7680
aaaaaaaaaa aaaaaaaa aaaaaaaa aaaaaaaa aacgaaaaaa aaagggggcg 7740
cc 7742
```

<210> 23

<211> 1674

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 4259314CB1

<400> 23

```
ggcggcggtg tctcaggcgg caatggaagg atccagacct gtggccgcccc atcaggggga 60
agaggcgctc tgttcttctt gggggactgg cagcacaaat aaaaatttgc ccatatgtc 120
aacagcatct gtggaatcgg atgatgcatt gtatagtcca cagaggtacg ttcttgaga 180
cacagcaatg cagaagatgg ccaagtccca tgtttcttta agtgggatgg gtggtcttgg 240
tttggaaatt gcaagaatc ttgttcttgc agggatttaag gcagttacaa ttcagtatac 300
agaaaaatgc caagcatggg atctaggaac caacttcttt ctcagtgaag atgatgttgt 360
taataagaga aacaggcctg aagctgtact taaacatat ttcaaacata atccatcgt 420
tcattgcaca tcatcttctg ttcttttcaa tgaagccaca gatctctct ttttagataa 480
atccagtggt gtagtattga ctgagatgaa acttccattg cagaagaaga tcaatgactt 540
ttgcctgtct cagtgccttc caattaagtt tatcagtgca gatgtacatg gaatttggc 600
aaggttattt tgtgatttgg gtgataaatt tgaagtttta gatacaacag gagaagaacc 660
aaagaaatct ttcatttcaa acataacgca agcaaatctt ggcattgtta cttgctctga 720
aaatcatctc cacaactcgg agacaggaaca attcctaaca tttcgagaaa ttaattgaa 780
gcaggtttta aatggatcta tacaacaact aacggtgata tggccatttt ctttagtat 840
tggtaacacc acagaactgg aacocatatt acatggaggg atagctgtcc aagttaagac 900
tcttaacaca gttttttttg atcactgga aggcagttta aaacatccaa agtgccatt 960
tgtggtttt agcaacctg aggcacctt agagattcac acagctatgc ttgccttga 1020
```

WO 02/097032

PCT/US02/11152

```

ccagtttccg gagaaataca gtcgcaagcc aaatgttggg tgccaacaag attcagaaga 1080
actgttgaaa ctagcaacat ctataagtg aaccttggaa gagaaggtga ctattgaaat 1140
ttatggtgtc cgaatatttt gtttgttaat acataagtg tctgtatatt agtattctct 1200
tatccctcac atccagcttt gctactctgt ggcattagac aacttttttc agtttggttc 1260
gttttaattt ataaacttat aaatagagt taacctacct tcatagggtt atttgaaaca 1320
ttaaataatt taattttgla agtacttag aaatgccatg tgacatatag taattaabat 1380
ttgctgtcat ttttttaact gttactatta ttctttatta ttctgttatg atttaactga 1440
ttgtttaacc cacttaacct tatatatgtt agctttctta tataagtcac ctgttttata 1500
tattatcatc tcatatctct ttatcscatt tctgacatca gatttagatt taaacatttc 1560
aactagaaaa tgggtttttt gtgtgtgtta tgatgccagt gcttgcctgt caacagtttt 1620
atttgagcag ctataaacat ggtgatgttt gtgcaattaa aaaaaaaaaa aggg 1674

```

```

<210> 24
<211> 3671
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 3660046CB1

```

```

<400> 24
agcagagggtg tgtacgggca ctgcttttaa actgggaagg aggaagacga ggcaggagc 60
tggagggtgca ccaaggtaga ttccagcag cgctagtcca gctgaacact ttccagcctt 120
gtttttcagc agctttgagg aaaaagtatg tgatcogtat gtgaacttt cattgtacgt 180
agcggatcag aatgaggaac ttgcttttgt ccagacaaaa acaattaaaa agacactgaa 240
ccccaaatgg aatgaagaat ttattttcag ggtaaaccca tctaataca gactcctatt 300
tgaagtattt gacgaaata gactgacac agacgaectt ctgggccagc tgaacgtgac 360
ccttagtcac ctccgacag aagatccaac catggagcga cctatataat ttaaggactt 420
tctctccaga ccaagaagtc ataatgtctg agttaaggga ttttttgcgt tgaanaatgc 480
ctatctgcca aaaaatggag gtcaagatga ngaaaacagt gaccagagg atgaactgga 540
gcattggtgg gaagtgtgtg actcaaatga ctggtctctc cagcaccaag aggaacttcc 600
tctctctctc ctgctctccg ggtgggaaga aaaaaggtag aatttaggac gaacttacta 660
tgtcaaccac aacaaccgga ccaactcagt gcacagacca agcttgatgg acgtgtcttc 720
ggagtcggag aataacatca gacagatcaa ccaggaggca gacacccggc gcttccgctc 780
ccgcaggcac atcagcgaag acttggagcc cagacccctg gaggccgggg atgtcccca 840
gccttgggag accatttcag aggaagtga tatcgttgga gactctctcg gtctgctctc 900
gccccacca ccggcctccc caggatctcg gacacgccct caggagctgt cagagggaat 960
aagcagaagg ctccagatca ctccagactc caatggggna cagttcagct ctttgattca 1020
aagagaaccc tctcaaggt tgaagtcagt cagtgtcacc gacgcagtg cagaacaggg 1080
ccatctacca ccgccagtg cccacagctg gaggagcgtt tcatcaactg tcaagggtgg 1140
tgaggaaaca acgcatcag tggcctatgt acataccac cggggtctgc ctccaggctg 1200
ggaagaaga aaagtgtcta aggggcgcac atactatgtc aatcataaca atcgaaacac 1260
aaactggact cgacctatca tgcagcttgc agaagatggt gcgtccggat cagccacaaa 1320
cagtaacac catctaact agcctcagat ccgcgggctt cgtagctca gctgcccaac 1380
agtaacttta tctgccccgc tggagggtgc caaggactca ccgtactgc gggtctgtaa 1440
agacacccct tccaaacac agtccccaca gccatccctc tacaactccc ccaaacacca 1500
acacaaagtc acacagagct tcttgcacac cggctgggna atgaggatag cgcacacagg 1560
ccggcccttc ttcattgatc ataacacaaa gactacaacc tgggaagatc cagctttgaa 1620
atttccagta catatgggtt caaagacata tttaaacccc aatgatctag ggcctttacc 1680
tccagatagg gaagagagaa ctcaacacaga tggagaatc ttctacataa atcaacaatc 1740
aaaaagaaca caatgggaag atcctcgggt ggagaatgta gcaataactg gccacagagt 1800
gcctactctc aggtattaca aaagaagta tgaattcttc cgaagaaggt tgaagaagca 1860
gaatgacatt ccaacacaaat ttgaatgaa acttcgcga gaaactgttc ttgaagactc 1920
ttacccgaga attatgggtg tcaagagagc agacttccgt aaggctcgac tgtggattga 1980
gtttgatggt gaaaagggtt tggattatgg aggagttgcc agagaatggt tcttctgatc 2040

```

WO 02/097032

PCT/US02/11152

```

ctcaaggaa atgtttaacc cttattatgg gttgtttgaa tattctgcta cggacaatta 2100
taccctacag ataaatccaa actctggatt gtgtaacgaa gatcacctct cttacttcaa 2160
gtttatttgt cyggtagctg gaatggcagt ttatcatggc aaactgttgg atggtttttt 2220
catcgcgcca ttttacaaga tgatgcttca caaaccata accctcatg atatggaatc 2280
tgttgatagt gaattattaca attccctaag atggattctt gaaatgacc caacagaatt 2340
ggacctcagg ttatcatag atgaagaact ttttgacag acacatcaac atgagctgaa 2400
aaatgttggg tcagaaatag ttgtccacca taagacaaca asggaatata tttatcttgt 2460
aatcaaatgg cgatttgtaa accgaatcca gaagcaaatg gctgcttita aagagggtat 2520
ctttgaacta ataccacagg atctcatcaa aatttttgat gaaatgaaac tagagcttct 2580
tatgtgtgga ctggggagatg ttgatgtgaa tgactggagg gaacatacaa agtataaaaa 2640
tggctaacag gcaaatcabc agtttatata gtggttttgg aaggctgttt taatgatgaa 2700
ttcagaaaaa agaataagat tacttcagtt tgcactcggc acatctcggg tgcctatgaa 2760
tggatbtgtc gaactatacg gtccaatgg accacagtca tttacagttg aacagtgggg 2820
tactcttgaa aagctgcccc gagctcatcc ctgttttaat cgcctggact tgcacacctt 2880
tgaatcattt gaagaattat gggataaact tcagatggca attgaaaaa cccagggtct 2940
tgatggagtt gattagatta caaataaaca tctgtagtgt ttttactgcc atagttttat 3000
aaccaaaatc ttgacttaaa attttccggg gaactactaa aatgtggcca ctgagtcttc 3060
ccagatcttg aagaaaaatc tataaaaagc atttgaaaga atagtacgac aacttatttt 3120
taataccttt taataaatgt gttgcattta cacagtgttt tcatctctgc tttagagtta 3180
gggtccctgc taaagccagg caccaccaca cctggcttta gagtccacac aataggatat 3240
aagtccctga tgaactaaat agtgaatttt gtcccttaaca tttactcttt gtatagtatc 3300
tgccagcgag ttttttctta aactactgag atgataactg tgaatatatt gtgatactgt 3360
tcatgtgtga aaagtctgat gcattttgag atggaanaact gaatttggg aaaaagaata 3420
ctttactatt gagtaaaact caatataatt agtgcactc gcagctattt attattttgt 3480
agacctgcct tatgcacctt actgcctaga tttttgggaa aaaactttgg aaagtgtgtt 3540
acctatattt ctagccaact aactcacaga aaaactgttt acttcttcac tttcgaagta 3600
tttggctttt gttaaatatg agttttacta aacagatggt tcataagaca tgtgaagcaa 3660
attcatattt g 3671

```

<210> 25

<211> 2038

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Ineye ID No: 3016416CB1

<400> 25

```

cgtttatttc cccaggaaat tttgaataaa gcccaaggcc cagaggggaa cccctttgag 60
gcccaaggga ggtccaaaat ttgaaactga gccaatccgc ggggacaccc tgaagtcca 120
caaaaaaaat ttacaaggca aattaattgg ggattcatgg cagacacctg tggttccgcg 180
tacttgggag gctgaggtgg gaggatcaac tgagcccgag aggtttgagtc ttgcagttag 240
gctgagttca caccactgta ctcgagcctt gatgacagaa tgagactgtc tcaaaaaaaa 300
aaaaatgtcc ttaagtccat gtggaccctc gactaggttt gtgcccctga cagccgtcct 360
ctgaggggca ttacagtggt gagactccag gtttaaatgg cctccacaga aatttcaact 420
acctgccttt gggtttgacc ctgtataacc cctttcttct ggaggtccct ttgggtggca 480
gtagatacgg gatttgggtg ctgacagctc tggggacaga tccagatccc aaatggcaga 540
gtctctacag attacaagcc aaataacttag cactatgtgc tgatcttcag gaagtccagt 600
tatatttcat aacaagtcac atggggataa tgaagggaatg gcttaaaatg ctctcagtaa 660
tattctctag tcatccctca gggctaggtt tgggttttag catggcgggg aaggagcag 720
agctgtgtgc agaggaagat gaagtcttgg cctgttcagg gtccctgacc tgatggcgac 780
ccatgttggg gtcttcatag tgacagacac cactgtaaaa gcagatccag gttgtgcaac 840
ctcacaagca ggtctctca ctacccggga tagatagact attggccgta cctgcatcca 900
cggcttgcca tggtttcgtt gtgggtggag gatattttcc tgtccctgg ctttgggttt 960
gccacgtggc ctgtctctgg ccttgggaatg aagcagaaac gaaaggctgc cagtcccgag 1020

```

WO 02/097032

PCT/US02/11152

```

ccacagctctg aagtcgcttt aggtgggttcc gggggggggg tgcgtccca cctccacca 1080
gaggggcttc tctgggtcag cagctgcttc ttcagctcc gcccaaaagg aacggagccc 1140
cctggcgagat cggcaggcct acagggagcc acagagcgca cgggctggac cagcgttcaa 1200
gcccaagcac aggcctcgca gaaccttgtt ccagcgcccg tttaggatgg ttgattagga 1260
cgctgtgcag tggcgttagc tcaccaatcc agtcgctgca cccgctcctt tattaggcta 1320
tagagccagc ggctccacca gggacctgat acacagtgcc gtttaataag gacatattg 1380
agctctcatg tcttaagcca gtggagaagt ccagggtcag tgtgggggct cggcgggggg 1440
ctgtggcccc catccgcatg gagcctcccc atggttcaca ggtctcagtc ttcggagcct 1500
tcggcctctg gagcccgacc agtcccaagg cggcgccgag accctcttct gaagcccatc 1560
ctctaaagcc tcggctccaa ccggtttccc ttcttcagcc tcaggatttt cactctcttc 1620
gaatgggggt ggccctcccc caatctctct agtcgcaaca gactctcctt cctccaggga 1680
cctcagagcc agagctgggg gagagggcct gaacctcggg gttaggttgg agcgtccctg 1740
tgaaggtgca gtcctgcttc ccacccccag gcgcggggcc tctcccaccc tcagcgccct 1800
gctcacctcc agctgaagat gccagggcac ctctgcttcc tccctgcccc ctctcgagta 1860
cggcgagctg tgcataaaag ggtttaatat aggtctttgc gggcgggggg actccacatc 1920
gtaatccagc taagttaga gaccaaggcg ggaggtatcc ttgaggccag gagtccaaa 1980
cagcctgggg caacaaagtg aggcccgctc ctgaaaaaaa aaaaaaaa aaaaaggt 2038

```

<210> 26

<211> 2235

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2133755CB1

<400> 26

```

tccagtgagc ggcggagccc ggcggcgccg gctggggccc gggcgggcgg ggtcggggc 60
tgagaggcgg gggggcgggg ggcggcgccc gggggcgccc catgtggagc ggcggcagct 120
cctccaccag cttgtgtgtg ggcgtgttgg tggctctacg gtgtcacacc tgcctgggta 180
tgtacggcat cgtctacacc cgcctgtgct ccggcgagcc caactgcacc cagccctacc 240
tgccggcgcc gccaagctg cagctgagct gtgtacacac gacgaggtcc cactgggtg 300
ctgagaacaa catcgacctg gtcttgaatg tggaaagact tgatgtggag tccaaatttg 360
aaaggacagt taatgtttct gtaccaaaga aaacggagaa caatgggagc ctgtatgct 420
acatctctct catcacgct ggggtctctg cgtggcacga cgggaagcag gtgcacctgg 480
tcagctctct gaccacctac atgggtcccc agccagaaga aatcaacctg ctcccgggg 540
agctctgata acagcagatc gagggcgaga agaagccgac gagtgccttg gatgagccag 600
tgtccacctg gcgaccggcg ctggcgctga acgtgatggc ggaacaattt gtctttgacg 660
ggtctctcct gctgcggat gtgcctcggg acatgaagat gatccagctg gggaaaaacg 720
tgcatctacc gccatccctg ttcatcgacc agctcagcaa ccgggtgaag gacctgatgg 780
tcataaaccc ctccaccacc gagctgcccc tcaccgtgtc ctacgacaa gctctcactg 840
ggcggtcgcg cttctggatc cacatgcagg acgcccgtga tccctcgag cagttcgggt 900
tttcagagaa agatgctgat gaggtgaaag gaatttttgt agatacaca ttatacttc 960
tggcgctgac cttctttgtc gcagcgttcc atcttctctt tgatttctg gcccttaaaa 1020
atgacatcag tttctggaa gagaagaaga gcatgatcgg catgtccacc aaggcagtg 1080
tctggcgctg cttcagcacc gtggtcatct tctgttctct gctggagag cagacagacc 1140
tgctgggtgt ggtcccgggg ggtgttggag ccggcattga cgtgtgaaa gtgaagaagg 1200
cattgaagat gactattttt tggagagggc tgatgcgcga atttcaagttt ggcacttaca 1260
gcgaatctga gaggaaaaac gagaggtacg atactcagcc catgaagtao ttgtcatacc 1320
tgctgtaccc tctctgtgtc gggggtgctg totattcact cctgaatata aatataaga 1380
gctggtatcc ctggttaata aacagcttcg tcaacggggg ctatgccttt ggtttctct 1440
tcagtctgcc ccagctcttt gtgaactaca agttgaagtc agtggacat ctgcccctga 1500
aggccttcaa ctacaaggct ttaaacacat tcattgatga cgtctttgac ttcactatca 1560
ccatgcccac gtctcccgcg ctggcctgct tccgggacga cgtggtgttt ctggtctacc 1620
tgtaccagcg gtggttttat cctgtggata aacgcagagt gaacagttt ggggagctct 1680

```

WO 02/097032

PCT/US02/11152

```

acgaggagaa ggccacgcgg gcgccccaca cggactgaag gccgcccggg ctgcgcgcag 1740
ccaggtgcaa cttgaattgt caatgagtat ttttgggaag atttgaggga attcctagac 1800
attgcgtttt ctgtgtttgcc aaaatccott cggacatttc tcagacatct ccaagttcc 1860
catcacgtca gatttgagc tggtagcgtc taggatgccc ccacgtgtga acatctgtct 1920
tggtoacaga gctgggtgct gccggtcacc ttgagctgtg gtgctctccg gccacagagt 1980
gtccgggggtt cggccatgtc ctacgcgcgg cagggtgtgg agcctcaca gccaaagggg 2040
ctgttggtatt tccatttcag gtggttttct aagtgtctct tatgtgaatt tcaaacacgt 2100
atggaattca ttccgcacgg actctgggat caaaggctct ttctctttt gtttgagagt 2160
tggttgtttt aaagcttaat gbatgtttct atttcaaat aaatttttct ggctgtggca 2220
aaaaaaaaaa aaaaa                                     2235

```

```

<210> 27
<211> 1851
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Ineyte ID No: 5259957CB1

```

```

<400> 27
ctcggaacta ctgcttgatt ctctgagaga tcccagcacc ctacaaactg agtccagatc 60
tgagttttcc cttgcagatt catcaagatg agcatcaggg ccccaaccag actcctggag 120
ctggcaaggc agaggctgct gagggaaccg gccctggcca tctccaccat ggaaggagctg 180
ccagggagc cttcccccac ctgtttcatg gaggccttca gcaggagacg ctgtgaacc 240
ctgaaaacaa tggtagcagg ctggccttcc accgcctccc ctctagggtc cctgatgaag 300
tgcctctatc tggagtcatt aaaaactgtg ctggaagggg ttgatgtgt gttgaacca 360
gaggtttgcc ccaggcagtc aaaaactcaa gtgctggact tgaggaatgt ggtgagaaac 420
ttctgcgaca tatttctcgg agctactgca tccctcccg aggcctctgag tcagaagcaa 480
acagacata actgtccagg gacaggcagg cagcagccat tcatgtgtgt catagacctt 540
tgcttcaaga acaggacact agatgaatgc ctacccacc tcttagagtg gggcaagcag 600
agaaaaggct tactgcatgt gtgttgcag gactgcagg tttttggaat gcccatccac 660
agtatcatag aggtcctgaa catggtggag cttgactgta tccaggaggt ggaagtgtgc 720
tgcctctcgg agctgtccac tctgtgtga tttgcctt accctggcca gatgaggaat 780
ctccgcaaac ttgttctctt caacatccgt gcactctgct gcattccccc agacaacaa 840
gggcaagtta ttgcctgatt caactctcag ttcctcaagc tggactatatt ccaaatctg 900
tcatgcaact cgtctctctt cctcgaaggc caactggacc agctgtctag gtgtctccag 960
gcctccttgg agatggtcgt tatgacagac tgcctgctgt cagagtcaga cttgaagcat 1020
ctctctcttg gcccgagcat ccgtcaaatc aaggagctgg accctgaggg tgctacgctg 1080
aoccatttca gccctgagcc cctcacagg ctgctggagg aagctgtggc caccctcgag 1140
accctggact tagaggactg tgggacatgt gatttccaac tcagcgccat cctgcctgtc 1200
ctgagcgcgt gctccagct cagcaccttc agcttctgtg ggaacctcat ctccatggct 1260
gcccttgaga accctgtcgg ccacacccgt gggctgagca agctaagcct ggaagtgtat 1320
cctgccccctc tggagagtta tgacaaccag ggtgtctctc gctgggggag atttgtgaa 1380
cttgggggtc agctgatgaa cacactgagg gacttaaggc agcccagat catgtgttct 1440
tgccacgtcc cctgccccctc ctgtggcatc agggcctcct atgacctgga gccagtcac 1500
tgccctctgt gaatgcctgc catcaggggt gatatttttc aagctttctt ctggtcatat 1560
cggagctgaa accctaggcca tgagtgcagt ttaaaggag caccagacca tggtttcaaa 1620
tgctcctca gtgtgaatgg gaaaggatg aggatgcagg aggggacagga ctgggggaaa 1680
agtgtacttg gagtggatgg gctctttaga gacctgtgt ccagagaatc agaaatggga 1740
atctgaattg ctagagtgag aatcagggag gagagacaca tgagaggggt acccctgcac 1800
agatggttgt aaagtaacag toagaaataa agggaaactg agtggaaga a 1851

```

```

<210> 28
<211> 1466
<212> DNA

```

WO 02/097032

PCT/US02/11152

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 55029783CB1

<400> 28

```
aaagacgcaa ggcgcgcgcg cccaaggetc agcgcgcgctg cgcaggatag cggccggtta 60
gcccagcgct cggcgccgga agcactggag ccccgagctc cgtgcgtgcg gcccgagatg 120
agcggggcgt gggacgtgct gcggcgctct agctggctta caggcgggcg gccgggtatg 180
tgtctctctg taagagtgct actgcgcgg ggttgabctg tgcctgccac tctcggttca 240
gacggtgagg tggcgctctg cagggaacgc ggggtggagt agaaggcgag ccggagacag 300
gcccggcgcc ccttcggagg ctagacggcc ccagcttcgc ggggacatg gaattcttgt 360
ggaccgagtg cggggccact ggcgaatcgc gcgggggtcc tgttcaacct gctggtgtcc 420
atctgcattg tgttctcaca caaatggalt tatgtgtacc acgggcttcc caaacatgag 480
cctgaccttg gtgcacttcg tggtaacctg gotgggcttg tatatctgcc agaagctgga 540
catctttgcc cccaaaagtc tgcgcgcctc caggctctcc ctctcggtcc taagcttctg 600
tggctttctg gtcttcacta acctttctct gcagaacaac accataggca cctatcagct 660
ggccnaggcc atgacccagc cgggtgatct agccatccag acctctctgt accgaaaaac 720
ctctccaccg agaatccagc tcacgtgatc tctataactc ttagggtgtaa tcttaaatc 780
ttattacgat gtgaagttaa atttctctgg aatgggtgtt gctgctcttg gtgttttagt 840
tactacctct tatcaagtgt ggttagggag caaacagcat gaattacaag tgaactcaat 900
gcagctgctg tactaccagg ctccgatgta atctgccatg ttgctgggtg ctgtgccttc 960
ctttgagcca gtgtttggag aaggaggaaat atttggctcc tggctagttt ctgcttttgt 1020
tttggtgctg ctatctggag taatagcttt catgggtgaac ttatcaattt attggatcat 1080
ttgggaacct tcacgtgtca cctataacct gttcggacac ttcaagttct gaattacttt 1140
attcgaggga tatgttttat ttaaggatcc actgtccatt aatcaggccc ttggcatttt 1200
atgtacatta ttggcatttc tgcctatcac caactttaag ctacgtgaac aggaaggaaag 1260
taggagtataa ctggcaccaac gtccctaatt gggtttttgt ggagaaaaa atgttgttcc 1320
aagaagataa aaaatattgt taagtgtgca agttattaaa aaaaaaaat tgggcaggc 1380
acggtggctc acgctgttaa tcccagcact ttgggaggcc aaggccagcg gatcacttga 1440
ggtcaggggg tcgagacagc ctgaca 1466
```

<210> 28

<211> 1049

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 8032202CB1

<400> 29

```
gtgcagcccc tcccacagc atgctggggg ctaattctga tgtcatcttt ctgcagaaaa 60
coattagacc atccctccag actgcacccc tcaaacgctg ctgcccaggc cccatctgac 120
actcttgaca totgcaggtc ccagacccta tgaatgtctc actctggagg ctctctatct 180
tctctgggtt gctggccttg cctttggcac cacacaagca gctctggcct gccctggccc 240
aagcccacag agacaacaaa tccacccctgg caagaattat tgcctagggc ctcataaagc 300
acaacccaga aagcogaatt cagaacatcc actttgggga cagactgaat gcttcagcac 360
aagtggcccc agggctgggt ggctgggtaa tcagcggcag gaaacccag cagcagcaag 420
agaagcagct caacatcacc aacattccag tggactgttg tgggatccag atactattcc 480
ataaggagtg gttctcggca aatatctcac ttgaatttga ccttgaattg agacctctct 540
tcgtatacaa catctgaaag atgtgtgcac atatagcatc cgttctggag tctctggctg 600
agaagaacga gtttggccgg aggggatctg tgataggcaa atgcgatgca gagccacga 660
gtgtccatgt ggccatctct actgagccta tcccacaaa gatgaatcag tttctctaca 720
acotcaaga gaatctgcaa aaagtctccc caccatctgt agaagtccag gtatgtctcc 780
```

WO 02/097032

PCT/US02/11152

```

tgatcgggtga aatcctcggg cagctggatg tgaacctgtt gaaaagcctc atagaacagg 840
aggctcgctca tgaaccaacc caccatgaaa ccagcccaac ctggtgcacg ccaggctgga 900
gagtcacccca gctgactttct gctgatcaga aggaagctcc acatcttgca accttaagtc 960
tcctctagag tggggctttct gctacccctaa aaactttacc ccaggctctg tggacatacc 1020
atcctctctct acataaact ctagctctg 1049

```

<210> 30

<211> 2520

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 6937367CB1

<400> 30

```

tgcccgccgg cccctccgcc toactcagcg gcgccaactga gagggaacgg cgccagccat 60
ggagcgacaca gcaggcaanag agctggccctt ggcacccgtg caggactggg gtgaagagac 120
cgaggacgggc cgggtgtaca gtgtctccctt gcggcggcag cgcagtcagc gcaagagccc 180
ggcggaggggc cccgggggaca gccaggtctcc cagccccatt gccaatacct tccctccacta 240
tcgaaaccagc aaggtgagggg tgcggaagggc agcgcgcctg gaggcggctgg tgggtagagtt 300
gggtgtttgga gacctgagcg aggaacccagc cttcatgccc gccctcctgg cccctacccg 360
gaacctttgta cccactgcct gctgtcgggg cttctgtctg ccaccaatgc caccgcccc 420
acotccccgg gtatagatca agaagacagc ggtacaagat ctgagcttca acaagaacct 480
gagggtcgtg gtgtcagtg cgggtcctct gctgcaggac caccctcagg atttccgaga 540
ccacctctgcc cattcggaac tgggcaggtgt ccgaaccttt ctgggctggg cggccccagg 600
gagtgctcag gctcaaaaag cagagaagct tctgggaagat tttttggagg aggtctagcg 660
agagcagagaa gaggagccgc ctacagtggt gacagagacct cccagagttg cccaaacttc 720
tgacccagac tottcagagg cctgcgcgga ggaagaggaa gggctcatgc ctcaaggtcc 780
ccagctcctg gacttcagcg tggacagaggt ggcggagcag ctgaacctca tagacttggg 840
gctctctctcc aaggtgagcg tctacagagt cttgggctcc gtgtggctgc agaggagccg 900
gocgggggct caggagccct ccccccactt scggccccacc gtggcccaat tcaacacctt 960
gaocggcctgt gtgtcgggtt cctgtctcgg agcaccgggc ttggcccgccc cgcagagggc 1020
gcagcggctg gagagtgga tccgcatcgc ccagcgctgc cgaagactgc ggaacttttc 1080
ctccttgccc gccatctctt cgcctctgca atctaaaccc atctaccggc toaagcgccg 1140
ctggggggca gtgagccggg aacccgtatc tactttcagg aaactttcgc agattttttc 1200
cgatgagaaa aaccacctca gcagcagaga gatctcttcc caggagggag ccaactgagg 1260
atcccaagaa gaggacaaca ccccaaggcag cctgcccctca aaacaccccc caggccctgt 1320
ccctacacctt ggcaccttcc ttacggacct ggttatgtct gacacagccc tgcgggatat 1380
gtttggagggg gatctcatla actttgagaa gaggaggaag gagtgggaga tcttggcccg 1440
catccagcag ctgcagaggg gctgtcagag ctacacccct agcccccaac cgcacatctt 1500
ggctgccttg catgcccaga accagctcac ccaggagcag agctaccggc tctccgggtt 1560
cattgagcca ccagctgcct cctgccccag ctccccacgc atccgacggc ggratcagct 1620
acccaagcgt ctacgtgcga agcttgcccc agagaaaagc tcatcaccta gtgggagttc 1680
ggggggcccc tcatccccca cctccagtgt gtccccagg tccccccct caagctctag 1740
aagcagagat gctcctgctg gcagtccccc ggccctctca gggccccagg gccccagcac 1800
caagctgccc ctgagcctgg acctgcccc ccccccggcc ttccgtttgc ctctgggcag 1860
ccctcgatcc cccctccggg cgcagcagag ctccggaggcc cgtgtcatcc gcgtcagcat 1920
cgacaatgac cccgggaacc tgtatcgaa catcttgctg accagtcagg acaaaagccc 1980
cagcgtggto cggcgagcct tgcagaagca caatgtgccc cagccctggg cctgtgacta 2040
tcagctcttt caagtccttc ctggggacgg ggtgtcctct attcctgaca atgccaagct 2100
ctctctatgcc atgagtcacg tgcacccag agacttcatg ctgcggcgga aagaggggac 2160
ccggaacact ctgtctgtct ccccaagctg aggcagccct gtctcttcca caagacacaa 2220
ctgctctggt caagcttgcc actctctctc tggaaagctg ccatcccca gttagagcca 2280
ctgtgtctgt tatccacgga ccaaccccca actgtagccc attggacccc atctcttttt 2340
ctgactctgt tggtagtaga tccatattcc aagacatca gcccatgggt ggctgggtga 2400

```


WO 02/097032

PCT/US02/11152

gagctcaatc ccacaaatgt agaaagaggt ggggcatgga tacgtcaaat cccctccctag 2460
 agaaatctta taaatgttag agacgcatac gaagtgcacac atgcggatga actatgcacat 2520

<210> 31
 <211> 4360
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3876510CB1

<400> 31
 gagatgacaa tagggagaat ggagaacgtg gaggtcttca ccgctgaggg caaaggaaagg 60
 ggtctgaagg ccaccaaggga gttctgggct gcagatatca tctttgctga gcgggcttat 120
 tccgacgtgg tttttgacag ccttgttaaat ttgtgtgccc acacctgctt caaaggagcag 180
 gagaagctcc atcgctgtgg gaagtgcagg ttgcccatt actgcgacog cactgcccag 240
 aaggatgctt ggctgaacca caagaatgaa tgttcggcca tcaagagata tgggaagggtg 300
 cccaatgaga acatcaggct ggccgcccgc atcatgtga ggggtgagag agaaggcacc 360
 gggctcacgg agggctgcct ggtgtccgtg gacgacttgc agaaccacgt ggaacacttt 420
 ggggaggggg agcagaaggga cctgcgggtg gacgtggaca cactcttga gtaactggcc 480
 ccgcagagcc agcagttcag catgcagtac atctgcgaca tcttcggagt gattaactgc 540
 aacggtttta ctcrcagtga tcaagagagg ctgcaggccg tggggctagg catcttccc 600
 aacctggccc tggtagacca tgaactgttg cccaactgta ctgtcatatt taaccaatggc 660
 aatcatgagg cagtgaatc catgtttcat accagatga gaattgagot ccgggccccta 720
 ggcaagatct cagaaggaga ggagctgact gtgtctata ttgaactcct caacgtttag 780
 gaagaagcca agaggcagct gaagaagcag tactactttg actgcacatg tgaacactgc 840
 cagaaaaaac tgaagatga cctcttcctg ggggtgaaag acaaccccaa gccctctcag 900
 gaagtggtga agagatgat acaattctcc aaggatcac tggaaaagat agacaaggct 960
 cgttcaggag gtttgtatca tgaggttgtt aaattatgoc gggagtgcct ggagaagcag 1020
 gagccagtgt ttgtgtacac caacatctac atgctgcgga tgctgagcat tgtttcggag 1080
 gtctcttctc acctccagcg ctttgaggag gccctgttct atgcaggag gatgtggac 1140
 ggctatatga agctctacca ccccaacaat gcccaactgg gcatggccgt gatgcggcca 1200
 gggctgacca actggcagcg tggtaacatt daggtggggc acgggatgat ctgcaaacgc 1260
 ttgcacattc tctgtgtgac acacgggccc toccaacca tcaataagga ctttagagcc 1320
 atgcgggtgc agacggagat ggagctacgc atgttccgcc agaaccgaat catgtactac 1380
 aagatgcggg aggtgcctt gaacaaccag cccatgcagg tcatggccga gccacgcaat 1440
 gagccatccc cagctctgtt coacaagaag caatgaggac tccccagtg aggaagggcg 1500
 atgtgtgtgt ggagctaggg agagactctg gaggtgtgtg gtctctgggg agacccctaa 1560
 tgaggaaagt gaggtaatgc ttaacattgt tgcgtgtaga atttactgoc ctgtgtttcc 1620
 cagagccatt ttggctcaat tcaagtctat tcaattcaag ttaactctag cccagcccag 1680
 atcaactcct cctacaataa ttattggatg ataggcccta gaaccaata aaggagctcc 1740
 aaatgtcgtt gggtagggaa gcaaaatgta gagaacatt taaagcacac tgtaataata 1800
 aatgcaatta taaactatat ggagggaggt gcagaggagg gaattgtctt ggtgtgtgat 1860
 gtgtgtgtgt gcagtggggg tatcacagag agtatgacat ctgagttgag ggtagcaggt 1920
 gccctggatc tcaggtggct gctcaccat ctgtgcaggt gtctctgggg ctgtgtgtct 1980
 caoctgtgtt ctgcagtaga cacaattggc tgagcaggat atgtgatact gbtgtgtgtg 2040
 tgtggagttt tgaagaaggg gctgtgtttt ggccacgtag gctctactca gagacctgaa 2100
 accacttcag aatgggtcat atgtcgaaag agctggctgg gggccttgcc caaaccaact 2160
 gaggtcttaa agtcggggga aaaaaagtct gggttccaac tagaattcta gaaatatttc 2220
 tagaacacac agagagggaa taagtccctc tatcaccctt attaccaago cttgtgtttc 2280
 cctgtgattt tagataatgt ctgatatatt tctggctatt tgcoctagtag gatttaaaaa 2340
 atattttcaa agtgaagctg agagagaatc ttggaaacac acatacctgt tgatcattgg 2400
 cccgtcagaa ttggcccttg ggggctttat ttggttacat gtgcctgggt ggtctttacc 2460
 agcttagact ctatcatggg ccccaatgaa gctccattct caatactgaa taattattac 2520
 ttcccttgtt gagttttctt ttctgtcatg cccctgggggc ttctgtcttt ctcaccagaa 2580

WO 02/097032

PCT/US02/11152

```
agaacatttg aatctggatt ctgtacacc tgggttagac cctgttcaga ggtgtggcca 2640
atttatcccg atctcctgga aggtctgtgt gatttccatc taagaaatga ggtctttgag 2700
aatcaaacag tcccaagatt agcctgttat cctgttatct actgagacct caaattttct 2760
accaatgttt tgggagatcc tggaaaagat cctctcagtt tgggggtgca ccaagacttc 2820
tacacaaccc aggaactacca ttgacctcag agctgtacc cactcttga agtaattga 2880
tcccaccagg tcccacgttt gttatctctg cctaaatgtt agcttctcca tctcaccac 2940
atgatgaact gctgtgtccc tctgagcact accaagtggc tgaaaactct gaaatgggc 3000
cacacttttg caaaatactt gtatctgaca cttaggtctt gtttgaagaa ttctcttct 3060
ggaggttttt acaagaagac tgatagtctt tcaagccccc acatccaggc cttagggaag 3120
gcactaaact tctcccaggg atctaacctg ctagtccaaa ttatcactct ttacctcca 3180
tataaaatgt ctcctccaaa cctttttccc ttcttgtcca ttgttatctg cttaagccact 3240
ggtcatctcc coatatctgt agtctttttt tccatctat ctctctaata ttgtttgtct 3300
ttaacaaact gtgtctgtgt tctgtgtccc tcttccctcc tcagaccact ggaatgcaag 3360
tctctctctc ctttgggaat tactctggat ccttccctcc gcttggacc cagacttttg 3420
ctccatctat tattgtctct cctcctggga tcttggacat ttgtccccc actggccttc 3480
tcaggtgcaa tcagtaaaaa tgcagagaac tcttggatct taactctcat gactgagttt 3540
tttttagttg tatagtctac atctgccttt cttaactttg cacttctctt tgaatccatt 3600
gcagatgac ttccactccc actccttccc taaaagggtc cttaaccaaga tcaaatctaa 3660
tgggtacact ttagtctcta tgtgatttgg ccttctgagt tcaatcatca ctcccagcca 3720
ttgtatttgg tgacccactt cctgtgtagt atcttctgat ctagtctctc aggttctctc 3780
gctggtctct ttctcttccc tgcctctgac atattgacat ttctctggagt tggttttgtc 3840
cttgatctat tctcatgcca ttctgcacac agtctctgca tgaactcagg cagacccttc 3900
atttaactgac caccttaggg ctgatgatt tcaaatctgt attcccagat ctgcaattg 3960
agctccagcc ccaactctcc tctggaggt tctgcaggcc cagcaaaact atcatgtcca 4020
aagtgaactt tttctctctt cctgtctctc ctctctgat ctgttcttct ttggaaaccc 4080
acccaagaac gtcaactctc coactcagatt gtgagctctt ggaaggcagg agctgtgtcc 4140
ttctatctat ctctctatcc ccagaaactt gcacagatcc tggaaatgtg taggtgtcca 4200
gtaaatgtgt gttgaataaa tgaatgaatt aatgaacaaa tgaatgaatt tgcctacttc 4260
aaggcaaaag aacctgaaa ctgtatttgg agttctctat ttatagcagt cagcaaatcc 4320
tattaaatcc ttgtgttttc caagcaaaaa aaaaaaaaaa 4360
```

<210> 32

<211> 3500

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 4900076CB1

<400> 32

```
atccccggca cgtggtctct ggtgagcgcg gccctccgcy ctccttggcc ccaggatgcc 60
ctctctctgt gggaaaggag ggtcgggaaa ggcgagcgt aggtccacct tctccaatcc 120
ctgctctctg ggaagggagc atctcttgag aaaggaagca cttctgtgct cccgagaaact 180
tctatccagg tcttggctgc agggaaacaa gctgggcttt ttataatbaa ggttggaaaga 240
agtccaccac ggcagcagaa ctccatcttg agatgaaata acatctacct ggaactctgg 300
cagaatttca aggcacacac tgggtgact ctggcgccat gatgttgcct tatcttccag 360
cactgggaga tcaatactgg gaagagattt tgcctccaaa gaatggggaa aatgtagaga 420
ctatgaagaa attgacccaa aatcatabag cgaaggctt gccttctaact gatactgact 480
gccccagaa aaaggaggga agggcccaaa tagtggtacc agttacatc agggatgtga 540
ctgtgatctt cacagaagca gaatgggaga gactgagttc agagcagagg aatctatata 600
aagaagtgtg gctggagaat tacagggaat ttctctctat ggcagaaaca aagccagaaa 660
tctacacctg ttctctctgc cttctgtctc tctctctgca gcagttctct agtcaaatg 720
taactcagat ctctctgggc ttatgtgcag aaaatcactt ccatccaggg aattctagcc 780
cagggcattg gaaacagcag gggcagcagt attccatgt aagctgttgg ttgaaaatg 840
cagaaggcca ggaagaggga ggtggtctca aacctggctc tgcagggaca gaggagagag 900
```

WO 02/097032

PCT/US02/11152

```
aaacotcaag ggcattcccc agcccactcc aaagacagtc agcaagtcct agaaaaggca 960
acatgggtggc agaaaacagag cccagctcag cccaaagacc aaacotgttg cagctagaca 1020
aagggtttgaa ggaattagaa acccttgagat ttggagcaat caactgtaga gagtatgaac 1080
cggacccataa cctgggaatca aactttatta caaacccgag gacctcttta gggaagaagc 1140
cctacatttg cagtgaattgt gggcgaaagt ttaaaagatg atcaaccttc atcagacacc 1200
atcgtatatac ctcgatggag aagccttatg tctgcagtga gtgcggcgga ggttttagcc 1260
agaagtcocaa cctcagcaga caccagagaa cacattcaga agagaagcct tatgttgcc 1320
gggagtggtgg gcaaaagcttt agaagtaagt cctctctcaa tagacatcag tggactcact 1380
cagagggagaa gccctatggt tgcagcagat gtggcgaggt ctttagccag aagtcacotc 1440
tcctcagaca ccagaggaca cactccgggt agaaacctca tgtgtgctg gagtgtggac 1500
gaagcttttg tgataagtca accctcagaa aacaccagag gatacactca gggggagagc 1560
cttatgttttg cagggtaggt gggcgaggct ttacgcagaa ctccagatctc atcaaacacc 1620
agaggacaca cttggatgag aagccttatg tttagcaggga gtgtggcgga ggtttttgtg 1680
acaggtcaac cctcactata cagcagcgga cgcactctgg agagaagcct tatgtgtgtg 1740
gtgagtggtgg ccagaggtttt agtcggaat cactctctct tttccaccag aggacacact 1800
cagggggagaa gcatttatgtc tgcaggaggt gtaggcgagg ttttagccag aagtcaaatc 1860
tcctcagaca ccagaggacg cactcaaatg agaagcctta tatttgcagg gaatgtgggc 1920
gaggtctttg tgacaagtca accctcaattg tacatgagag gacacactca ggagagaagc 1980
cttacgtgtg cagtgaattgt ggcggaggct tttagccgaa atcactctctc cttgtccacc 2040
agaaggacaca ctccaggggag aagcattatg ttgttaggga gtgtggcgga ggttttagtc 2100
atagtcnaaa tctcactaga caccagagga caccactgac ggagaacct gtgtatgcag 2160
gggtcatgaa caagacctga gtgacagtc aagcctcatg ttacccocaga gagacacatg 2220
gggagtagac cctgtgtaca cagattgtga gtgaagtcc agagatgtgt cagccottat 2280
caggcatggg agggacacgt tcagagagga agccttatga gtatagagta cgggcaactg 2340
tagcctcag tcggccttga gcctgcacaa aaggacacac ttaggagaga agtttatgtg 2400
tagggactgt gggaaagctt tagcaataat caacatttac cagacatcca atgacagcct 2460
caggggaag caccctgttc tggggaggtg tggggagcat cagtataaga atggacactc 2520
agggcagag tgccctcag gaagaggtgc ttgttttgta ggtatgtatg gcaagcttt 2580
tgtgatcaca caccacaggg agaattctga tgtggggaca ctgtggagct ctgccagat 2640
gaccttttca gggtaaacac cccagctctc tgagagaaca gtgttgtac tggcagagat 2700
gcattocaga gatgcactcc gctctggaac tctctctcag ccacaggag ctgcactgac 2760
cacaggcgca atgcaccttt gccggggtag cttctggccc caacotttga ctcaacggg 2820
acaaotcag aagtcatto cagatccaga gatcccatc gaactgaag abcatgggt 2880
tcagacacaa ttgcaggtca gcttcttctc ctgcccagtc ctgcccac tccccagtga 2940
atctcctaatt ttctgtctcg ttgtctgtcc agataattga ttctaagaca tgttaggcat 3000
ataaggagtg tagataagga ttacagcatg agtaccoccc cagtaagccc cagagtatat 3060
tgaaatagaaa ttctgcactg gtggggagaa tggcaagga cttaggaaaa agtctctatc 3120
aaagaacagc ttttttggga caagctttac atgggtgggg aggggaatat tgaacacat 3180
tagcaataag ttaaacctca tctgtacta aaggaggaca cactcaggga gaagacotct 3240
gtgggcaggg cttgtgggtg gagcttcatc cagatgtcac tcttcaactg acttaggag 3300
acagttcttg taccocaggt cactacctca ctacactctg agggattttc aggaatatgc 3360
ttgactcccc catgtactct gtatgtgagc gaagatggca gtaactgtta aatagcatt 3420
cttttctact tcttggaaat agatgaaata aaaaagcagg cttttattta gtaatcaaaa 3480
aaaaaaaaaa aaaaaaaaaa 3500
```

<210> 33
<211> 1366
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1543848CB1

<400> 33
ttcggagggt cgcagcgagg tagatogcaa tacagggcct tgaataatga gaatttcctg 60

WO 02/097032

PCT/US02/11152

```

caaggccac acgctgcaa gggaaacggg ccggaggaa ttaatttcc cggggtgaa 120
gagaggctcg gctaattcgg cggccccccc tttttttttt tttttttttt tttttcggat 180
cgagcccttg ggcgggtggt gaggtggtaa ccgtgatagt agcagctccg ggggcagcaa 240
cagcgactac gagggatggc ggcggctgca gcaggaaact caacatocca gaggttttcc 300
cagagcttct cggatgccct aatcgacgag gacccccagg cggcgttaga ggagctgact 360
aaggctttgg aacagaacc agatgatgca cagtattatt gtcaaagagc ttattgtcac 420
attctctctg ggaattactg tgtgctgttt gctgatgcaa agaagtcctt agaactcaat 480
ccaaataatt ccactgctat gctgagaaaa ggaatatgtg aataccatga aaaaaactat 540
gctgctgccc tagaaacttt tacagaagga caaaaattag atatagagac ggggtttcat 600
cgtgttgccc aggtctggtc ccaactcttg aacctcaagg atccacttgc ctgggactcc 660
caaatgtctg gattacacgg tgcagatgct aatttcagtg tctggattaa aaggtgtcaa 720
gaagctcaga atggctcaga atctgaggtg tggactccat agtcaaaaat caagtatgac 780
tggtatccaa cagaatctca agtagtcatt acacttatga tcaagaatgt tcagaagaat 840
gagttaaatg tggaaatttc agaaaaagag ttgtctgtct tggttaaact tcttcttggg 900
gaagattaca atttgaactt ggaacttctt catctcataa taccagaaca gagcacgttt 960
aaagtacttt caacaaagat tgaattataa ctgaaaaagg cagaggctgt gagatgggaa 1020
aagctagagg ggcaggagga tgtgcctacg ccaaaaacat tctgagcaga tgtaaagaac 1080
ctatatccat catcatctcc ttatacaaga aattgggata aattggttgg tgagatcaaa 1140
gaagaagaaa agaattgaaa gttggagggg gatgcagctt taacagattt atttccagct 1200
atctattcag atgggttctga tgaagtgaag cgtgccatga acaaatcctt tatggagtcg 1260
ggtggtagcg ttttgagtac caactggctc gntgtaggta aagggaaggt tgaattcaat 1320
cctctgatg atattggaatg gaaaaagtac taataaattt aaattc 1366

```

<210> 34

<211> 4524

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 6254070CB1

<400> 34

```

gtggccgcag cgggttcttg agtgaattac ccaggaggga ctgagcacag caccactag 60
aggggggcca ggggtgcggg actcgagcga gcaggaaagg ggcagccctt ggcaccaggg 120
ctttgactca acagaattga gacacgtttg taatcgctgg cgtgccccgc gcaccagatc 180
ccagcgaaat cagatttctc ggtgaggttg cgtgggtgga ttaatttgga aaaaagaaat 240
gcctatatct tgcctacaaa aaactcacgg agggagaagcg cagtcaatca acagttaact 300
taagagtccc cggatgtctc cgttgtttaa acttgtatgc ttgaaaatta tctgaggggc 360
aatnaacatc tgcctcttcc ttccctctcc agaaagtcct tggaaatatta agccccaggg 420
ttgcttgggg atggctggaa gtgcaatgtc ttccaaagttc ttctcagtgg ctttggccat 480
atttttctcc ttgcgccagg ttgttaattga agccaatctc tgggtggtgc taggtatgaa 540
taaccctgtt cagatgtcag aagtatatat tataggagca cagcctctct gcagccaact 600
ggcaggactt tctcaaggac agaagaaaact gtgccacttg tatcagggac acatgcagta 660
catcgagaga ggcgcgaaga caggcatcaa agaattgccg tatcaattcc gacatcgaa 720
gtggaaactg agcaactgtg ataacacctc tgtttttggc aggggtgatgc agtatggcag 780
ccgcgagacg gccttcacat acgcggtgag gcgcagcagg gtcggtgaaog ccattgagcg 840
ggcgtgcgcg gagggcgagc tgtccacctg cggctgcagc cgcgcgcgcg gccccaaggg 900
cctgcacggg gactggtctc gggcgcgctg cggcgacaac atcgactatg gctaccgctt 960
tgcacaaggg ttctgtgagc cccgcgagcg ggagcgcatc cagcgcaagg gctcctacga 1020
gagtgctcgc atcctcatga acctgcacaa caacagggcc gcccgaggga cgggtgacaa 1080
cctggctgat gtggcctgca agtgcacatg ggtgtccggc tcatgtagcc tgaagacatg 1140
ctggctgcag ctggcagact tccgcagggt ggtgtatgcc ctgaaggaga agtacgacag 1200
cgcggcgccc atgcggctoa acagccgggg caagtgtgta caggtcaaca gccgctcaa 1260
ctgcgccacc acacaagacc tggctcaaat cgcacccagc cctgactact gcgtgcgcaa 1320
tgagagcacc ggctcgctgg gcacgcaggg ccgctgtgac aacaagacgt cggagggcat 1380

```

WO 02/097032

PCT/US02/11152

```

ggatggctgc gagctcatgt getgcggcgc tggetacgac cagttcaaga cagtgcagac 1440
ggagcggctgc cactgcaagt tccactggtg ctgctacgtc aagtgcacga agtgcacgga 1500
gatcgtggac cagttttrgt gcaagttagt gggtccaccc agcactcagc cccgtcccca 1560
ggaccocgctt attatagaaa gtacagtgat tctgggtttt ggtttttaga aatatttttt 1620
atttttcccc aagaattgca accggaacca tttttttttc tgtttaccac taagaactct 1680
gtggttttatt attaatatta taattattat ttggcaataa tgggggtggg aaccaagaaa 1740
aatattttatt ttgtggatct ttgaaaaggt aatcacagac ttcttttget agtatagaat 1800
gaagggggaaa ataacacata ccttaactta gctgtgtggg acatgggtaca catccagaa 1860
gtaaaagaaat acattttttt tttctcaaat atgcatcatc atgggatggg taggttccag 1920
ttgaaagagg gtgttagaaa tctattccaa attcagcttc tatgacccaa atgagttgta 1980
aattctctgg tgcaagataa aaggtcttgg gaaaaacaaa caaaacaaaa caaacctccc 2040
ttccccagca gggctgctag ctgtctttct gcattttcaa aatgataaatt tacaatggaa 2100
ggacaagaat gtcataattc caaggaaaaa aggttatcca catgtctcat tctctccaaa 2160
tattccattt gcagacagac cgtcatattc taatagctca tgaattttgg gcagcaggga 2220
ggaaagtccc cagaatttaa aaattttaa actcttatgt caagatgttg attgaagct 2280
gttataagaa ttgggattcc agatttgtaa aaagaccccc aatgattctg gacactagat 2340
ttttgttttg gggagggttg ctgaaacata aatgaatat cctgtatttt cttagggata 2400
cttggttagt aaattataat agtagaata atacatgaat cccattccaa ggtttctcag 2460
cccaagcaac aaggttaatt cgtgccattc agcactgcac cagagcagac aacctattg 2520
aggaanaaaa gtgaaatcca ccttccctct cactctgagc cctctctgat tctccctgt 2580
tgtgagtga tgcggccac gtttccaaac ggcagctcca ctgggtccc tttggttga 2640
ggacaggaaa tgaacacata ggaactctgc ttgaaaaa gttcactaat tagggatttt 2700
gtttctctaa aacttttatt ttgaggagca gtatgtttct atgttttaat gacagaactt 2760
ggcataagga attcacagag gtgttgacgc gtatcactgt tatgactctg tgtttagatt 2820
atccactcat gcttctctta ttgtactgca ggtgtacett aaaaactgtt ccagtgtact 2880
tgaacagtgt catttataag gggggaatgt tgggttaagt gtgctgata tctcaagct 2940
ttttgtcaat aacatataa tatatataa tatatataa taaatataa aatatactc 3000
attgcagcca gtgatttaga ttacagctt actctggggt tatctctctg tctagagcat 3060
tgttgtctct cactgcagtc cagttgggat tattccaaaa gttttttgag tcttgagct 3120
ggcctgtggc cccgtctgta tcatccctg agcagcaga agcaacctcg tttctaggga 3180
agaagcctga gttctgact actgaaatgc gtgttgggtt gaagatatct ttttttttt 3240
tctgctccac cctttgtct ccaacctcca ttctgttcca cttgtgggag agggcattac 3300
ttgttctgta tagacatgga cgttaagaga tattcaaaa tcaagaagcat cagcaatgt 3360
tctctttct tagttcatcc tgcagaatgg aacctcagc ctattagaaa tgacagtact 3420
tattaaatga gtccctaagg aatattcagc cactacata gatagctttt ttttttttt 3480
ttttttttta ataaggacac ctctttccaa acaggccatc aaatatgttc ttatctcaga 3540
cttaccgtgt tttaaaagtt tggaaagata cacatctttt cataccccc cttaggaggt 3600
tgggttttca tatcacctca gcaactgtg gctcttaatt tattgcataa tgatatccac 3660
atcagccaac tctggctctt taatttattg cataatgata ttacatccc ctcagtgtca 3720
gtgaattgtg agcaaaagat cttgaaagca aaagcacta attagttaa aatgtcactt 3780
ttttgttttt tattatacaa aaacctgaa gtactttttt tatttgttaa atcagattgt 3840
tcttttttag tgaactcatg tlatgaagag agttgagttt aacaactcta gcttttaaaa 3900
gaactattt aatgtaaat attctacatg tcatcagat attatgata tcttttagcc 3960
tttattctgt acttttaag tacaatttct tgttttgtg gatttgtata ttctactgt 4020
tataaaaaa acatcgaaa ggcattatgc aaatggaaga tagaatataa aataaaact 4080
tacttgtata ttgttaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaa 4140
taaaaaaaa aaaaaaaa aaaaaaaa aaaaaaaa aaaaaaaa aaaaaaaa 4200
taagcggcgc gcgcgcocca gtggggggca gccacagggc tcccttttaa gaacttttag 4260
aaagcggcgc gagagaacca ggggacacta ggtgagatgc agcggggaaa aaagtgttcc 4320
ccgggccaaa tgcaccaaca tbatctgaga agtgaagttt gaaaacgcca caaaattccc 4380
acagaataac gcacagggag gggaaacaaa caagaacaga agaggaacac agacagagc 4440
ggaaagagcc aagaaggagc ggaaaaaaac cgcaaaagga cccgcccgcg cggcagcgg 4500
accgccagag acaggacaca cgtc 4524

```

<210> 35

<211> 1157

WO 02/097032

PCT/US02/11152

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1289839CB1

<400> 35
tggagttgga cctggagaaa agtcaagtca taagtcaaga aagattgggc cctactactg 60
gaatgcagga aaaaatggag gaggatgga gaggtttga aaggcagcca caggsgttct 120
gggagagggg aggcattcta agtggcagta acagcttcag caaagtccca aaggtggaaa 180
agtgcaggac agtccaggc ataagccagt gcactaagcc cactcttctg cccacagtc 240
caggtggagg ccgacagagg cccaggggcaa gcagaggcag caatggtttg tctgacggt 300
ggctgagccc ccagcccttg gaatatgcag cccgggggag cccagacag cggcaaggac 360
gaggtggcgg agtggggcgg gaggcatggc ctccacctac cgggtggcgg tggggggggc 420
gcgaggtgtg ggcaagagtg ccatcgtgag ccagttcttg tacaacaggt taagcgaggt 480
ctggttcccc accaacgccc gccgccttta cctgcttgct gtgctcatga aaggccacgt 540
gacagacccc cagatccctg actttccacc catcagcgcc ttccctgtca atacgctcca 600
ggagtgggga gacacctgct gcaggggaat ccggagtgct cagcgctaca tctgtgtcta 660
cgacatctgc tgccttgaca gctttgagta cgtcaagacc atccgcacg agatcctgga 720
gacgaggggt atcggaacct cagagacgcc catcatcctc gtgggcaaca agcgggacct 780
gacgagggga cgcgtgatcc cgcgttgaaa cgtgtcgccc ctggtaagca agacctggaa 840
gtgctgctac gtggaatgct cggccaaagta caactggcac atcctgctgc tcttcagcga 900
gtgctgcaag agcgtggtgt gcgcccgttg caagcagctg cagcgtgccc tgcgcttcca 960
gggcgcgctg cgcgcgaacc gctgcgcat catgtgacgc ctgcgcgccc ctggggctgc 1020
acgggcactg gccgagcgga gggcggggcc gtaactcggg gctggggcgg ggagcgggcg 1080
gcaaatggaa ctgtgacggt cccggcctga ggcacctgca gccacgcacc tcccgtagg 1140
aagcagagcg cgagagg 1157

<210> 36
<211> 1419
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5565648CB1

<400> 36
ggacgcctgc tcagtgcggc cggcggggc aacctatgc tggcgtaate gggttcctcc 60
gagccgcctg aggaactggt ccggcgggct ggtgaggaat ggagccggta ggcgtgctgc 120
gcgagtgccc cggctcctcc gttagacccc ggagcacctt cgtgttgagt aacctggcgg 180
aggtgtggtg gcgtgtgctc accttccctg ccggcaaggc gttgctgagg gtggcctgag 240
tgtgcgctt atggagggag tgtgtgcgca gagtattgag gcccatcgg agcgtaacct 300
ggatctccgc aggcctggcg gaggccggcc acctggaggg gcattgcttg gttcgctggt 360
tagcagagga gcttgagaat gttgcctatc taccacatac agttctttac atggctgatt 420
cagaaaattt cattagctgt gaagagtgtc gtggccatac gagagcaagg aaanaacta 480
gtatggaaac agcacttgcc cttgagaagc tattcccaa acaatgcaca gtctctggga 540
tttgtgcccc aggaattgta gtgactccaa tgggatcagg tagcaatgca cctcaggaaa 600
tagcaattgg agaattggt ttgtccttat tattccctca aattgaagga ataaaaatac 660
aacctcttca ttttattaa gattccaaaga atttaacatt agaaagacat caactcactg 720
aagtgggtct tttagataac cctgaaactc gtgtggtcct gttcttgggt tataattgct 780
gtacggtggg aggcagtaat tatctgagc aagtgtcag cactttcagt gatagtaata 840
tcctctggc tggagggcag gtggcaacc tgtcatcact gacttttgaa aagaacctcc 900
tggatattga tgcctgggt gtggtggac tgtcatttag tggacccga atccagagt 960
ccactgtgct cctcaacgag gacgtcagtg atgagaagac tgcctgaggt gcgatgcagc 1020

WO 02/097032

PCT/US02/11152

```

gctcaaaagc ggccaacatt ccagagcata acaccattgg cttcatgttt gcatgcttg 1080
gcagggggctt tcagtattac agagccaagg ggaatgttga gctgatgca tttagaaagt 1140
tttttccatg tttccctta ttccgtctct ttgaaatgg aaaaattgga tgtgatcgga 1200
tagtcaactgg gaactttata ttgaggaat gtaatgaggt aaaagatgat gatctgttct 1260
atagctatcc aacaataatg gcaactcata atctgggggtc atctaaataa taattaaagt 1320
ggctttcata atatgtaact tttaggttct gctttttca gaaaatggaa aottgggcaa 1380
tgtgtatttt caacaaaaat accttagata tatctttt 1418

```

```

<210> 37
<211> 4113
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Inocyte ID No: 2764456CB1

```

```

<400> 37
tgctttatta accactctag gtattattct aacactttat gtgcttatta nattatcccg 60
acataattaga tttgctgaat tgcagtcocat atttcgtaga tatgatagct gaagcacgga 120
atcattaggt aacttgccca nagtggcatt cagactcoat aaatggctta gtagtbaaac 180
tcagttttat tccctgggac gccctctctg ctcttcagca cttgaagtcc aggcagcgag 240
agttagcatg gggccaggct gcccccctgg ggccgggttga agacagggtg agtctcttga 300
tattcaggaa atcatcgccg acccagtcac cagcgttcgg gacctctctg cagcgggacc 360
gacggaaacc ggagcaggcg acagggcgca gaagcgggat gtacttctgt tggggcgccg 420
actccaggga gctgcagcgc cggaggacgg cgggcagccc cggggctctg ctactgcagg 480
cggccagcgg ggagcgccac tctctctctg tctgtaccac ccacagggtc ctctctctg 540
gagacaacag cagggtctag ctgggcgcga ggggcgccga gcgcggggag ctgcagaaac 600
caattcagcg attggaaccc ctaattgttg atctctgtgag ctgcggggaag ggcactccc 660
tggtctctg ccacaaagga aggtctctcc catggggagc tggttctgaa ggcagctgg 720
ggattggaga attcaaggaa ataagtttca accctaagaa aataatgact ctgaatgata 780
taaaaaaat acaagtttcc tctggacact accctccct ggcattatca aaagatagcc 840
aagtgcttcc gtggggaaag aacagccatg ggcagctggg cttgggggag gagtccccc 900
cccaagccag cccgcagagg gtgaggtccc tggaggggat cccactggct caggtagctg 960
ccggaggggg tcacagcttt gccctgtctc tctgtgggag ttctgtttgc tggggaagta 1020
acagtgccgg gAgctggcc ctccagtgcc gtaatgtccc agtgcaagc aacaagctc 1080
tctcagctgg tgcactgaag aatctaggtg tggtttatat cagctgtggt gatgcaacaa 1140
ctgcggtgct taccagggac gggaaagtgt tcacatttgg agacaatgc tctggacagc 1200
tgggtatacag ccccaactcc gagaaagagc gtccacaact tgtggaaaga attgatggcc 1260
tagtttccga gatagattgt ggaagtctc acaccctggc atatgtgcac cccactggtc 1320
agggtgtatc ttttggctat ggaaccaagt acacaagcna gccaaactcat ccggagggcc 1380
tgacagagaa ctttgacatt agctgcctga tttctgtctg agactctgtg gatgttcaag 1440
tcaaacacat ttttgcctga acatatgcca acttctgtac aactcatcag gatactagtt 1500
ccacacgtgc tcccgggaaa accctgccag aataagccc aattagccag tccatggcag 1560
aaaaatggat agcagtgaag agaagaagta ctgaacatga aatggctaaa agtgaattta 1620
gaatgatatt ttcactcctc gcttgtctga ctgcaagtct ttaaaagaaa agaggaactg 1680
gagaaacgac ttccattgat gtggacttag aaatggcaag agataccttc aagaagttta 1740
caaaaaggga atggatttct tccatgataa ctacgtgtct cgaggatgat ctgctcagag 1800
ctcttccatg ccatcttcca caaccaagag ctttatcagt tttctctctg ctcccagaat 1860
gtctctgtat gcatgattct aagaactgga agaaactggg ggttccattt gcaaaagcctg 1920
tgtcttctct gaatacgtg atccagatgc tttaagcagc catcatctct cagctgcttc 2040
atcagactaa aaccgaacag gatcactgta atgttaaacg tcttttagga atgataaag 2100
aactgcataa gghaaacaaa gctaaactgt gactaccaga aaatactttc aacataaag 2160
aactctccaa cttattaaac ttttatatag atagaggaa acagctcttt cgggataacc 2220
acctgatacc tgcagaaccc cccagctctg ttattttcag tgattttcca ttactcttta 2280

```

WO 02/097032

PCT/US02/11152

```

attcgctatc caaaattaaa ttattgcaag ctgattcaca tataaagatg cagatgtcag 2340
aaaaagaaagc atacatgctt atgcatgaaa caattotgca aaaaaaggat gaatttctc 2400
catcacccag atttatactt agagtcagac gaagtcgctt gggttaagat gctctgcgtc 2460
aattaaagtca agctgaagct actgactctt gcaaaagtatt agtggttgaa ttattaatg 2520
aaattttgctc tgaatctgga ggggttagtt cagagttctt ccactgtatg tttgaagaga 2580
tgaccaagcc agaatacgga atgttcatgt atcctgaaat gggttcctgc atgtgggttc 2640
ctgccaagcc taaacctgag aagaaaagat atttctctt tggatgcctg tgtggactct 2700
ccttattcaa tttaaatggt gctaaccttc ctttccact ggctctgtat aaaaaacttc 2760
tggacaaaaa gcaatcattg gaagatttaa aagaactcag tccctgggtg gggaagagtt 2820
tgcaagaagt tctagatgat gctgcgatg acattggaga tgcgctctgc atacgtttt 2880
ctatcacact ggacaaaaat gatgttgact taattccaaa tgggatctcc atacgtgttg 2940
accaaaccaa caagagagac tatgtttcta agtatattga ttacatttcc aacgtctctg 3000
taaaagcagt ttatggagaa ttccagagag gattttatg agtctgtgag aaggagatac 3060
ttagacattt ctaccctgaa gaactaatga cagcaatcat tggaaactct gatlatgact 3120
ggaaacagtt tgaacagaaat tcaaaagtat agcaaggata ccaaaaatca cctcctacta 3180
ttcagttggt ttgggaaggct ttccacaaac taaccttgga tgaagaaaga aattctctct 3240
tttctcttac aggaactgat aggtctgatg caagaggcat acagaaaatg gaaatagtat 3300
ttcgtctgct tgaaaacttc agtgaagag atcacccaac atcaataact tgtcataata 3360
ttctctccct ccttaagtat tctacaatgg aaagaatgga ggaagcactt caagtaccca 3420
tcaacaacaa cagaggattt gtctcaccca tgcctcacaca gtcatataca cctctgagag 3480
actcagggtg ggcctttctca cacttggaac cttctgttct tccctacacc taaaataac 3540
aagagattaa tgaatagttg ttagaagtat ttgagggaga gatggggaga atggggagat 3600
gatgatgatg gtcaaaaggtt gcaaaatctc acacaagact gaggcaggag aatagggtac 3660
agagataggg atctaaggat gacttggaac cactccctgg cactgaagag tctgaacact 3720
ggcctgtgat tggctcattc caggaccttc atttgcataa ggtatcaaac cacatcagcc 3780
tctgatggc catgggcccag acctgcactc tggccaatga ttgggttcatt ccaggacatt 3840
ctttgcata aggaagtcaaa caacaccagt ctgggatttg ctgtgagcca attcaactca 3900
gtctctaatt ggctgtgagt cagtctttaa tttaactagg gtgtaaacat caagaaacct 3960
ctacaggggt ctttaagccc agaagatttt gctaccaggg ctcttgagcc actgctcta 4020
gcccactccc acctgtgga atgtacttcc acttttctg ctctactgcc ttgtgctcca 4080
ataaatccac tcttccaca ccaaaaaaaa aaa 4113

```

<210> 38

<211> 7058

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5734806CB1

<400> 38

```

cgttcogtta gggcggtggg gttggctgca gtggcagtc tttctcttct gctcacgggg 60
acctgctcag gctggaggcc agccagctct tgcggccacc tgggtcgga tggggcgcca 120
ggacggcgcg cagtgccact tgcacatcga gatcaaccgg gagccgggtg gtcgcattat 180
gtttcagctc ttctcagaca tatgtccaaa aacatgcaaa aacttctctt gcttggtctc 240
aggagagaaa ggccttgggg aacacactgg gaagaagtta tgttataaag gttctactgt 300
ccatcgtgtg gttaaaaact ttatgattca ggggtgggac ttcatgaaag gtaatggaaa 360
aggtagagaa tcaatttatg gtggatattt taaagatgaa aactttatct tcaaacatga 420
cagagcgttc cttttatcaa tggcaaatcg agggaaacat accaatgggt ccaagttttt 480
catacaaca aagcctgctc cacacctgga tggggtgcat gtagtctttg gactggttat 540
ttctggtttt gaagtaatcg aacaaattga aaactctgaag accgatgctg caagcagacc 600
atatgcagat gtgcgagtta ttgactgtgg agtacttgcc acaaaatcaa taaagagatt 660
ttttgagaaa aaaaggaaga aacaaactca ttcaagaagg tcggattcct cttccaattc 720
ctcctctctc tcagaatcat cttcagaag tgaacttgaa catgagagaa gcagaaggag 780
gaacataaag agggggccaa aagttaaac ttctaaaaag agggcaagg aagcaaggag 840

```


WO 02/097032

PCT/US02/11152

```

ttcagaagag ccaagggaata aacatgcaat gaacccaaaa ggtcactctg agaggagtga 900
taaccaatgaa aaaagggtcag ttgattccag tgctaaaagg gaaaaacctg tggctcgcgc 960
agaagagagatt cctccagtcg ctgagaacccg atttttactg agaagagata tgcctgttgt 1020
taactgcagaa cctgaaccaa ttctctgagt tgaccccatg gtaagtgate agaaaccttc 1080
tgtatcaaaag tctggacgga agattaaagg aaggggcaca attcgctatc acacaccttc 1140
aagatcaaga tctgtttctg agtcaagatga tgaatgcacg agtgaacttc ctctcactg 1200
gaagagagaa atgcagagatg taagagcata tagaccacct agtggagaaa aatggagtaa 1260
aggagataag ttaagtgaac cctgttcaag cagatgggat gaagaagctg tgtctcagag 1320
atccagatca tggctctata atggatatta ttacagacct agtaccgcaa gacactctgg 1380
ccacccataaa aaacgcagaa aagaaaaaaa ggtaagcat aaaaagaaag ggaagaagca 1440
gaaacctctg agagacaca aacaaacaaa gaagagaagg attcttatac cgtctgacat 1500
agaattctca aaattctcca ctgagaagat gaattctctt tgtgatagag aaaggagtgc 1560
tggttctctc tcattgtcat ctcatcactc atcaaaagag gactgggtca aatctgataa 1620
ggatgtccag agctctttaa cccattccag cagagactca tcaagatcaa aatctcactc 1680
acagtcttat tctagaggaa gctcaagatc aaggactcgg tcaaatgctt catcacatc 1740
tcgaagttaga tcaaatgcca gatctagtgc aagttctggg caccgaagaa gactatcaaa 1800
atccccaaga aaacagcttt ctcatgtaag tgaataaaaa ccagttanaa cagaaccttt 1860
aagagcaaac atggcacaata atgaaatgt agtagtaca ccagtgttag cagaaaaat 1920
tctctataa ccaactgagt ccagtccccc cctctcaaga tggagcctg gacagaaaac 1980
tgggaagccc tcttatgagc gaattcagga aatgaagctt aaaaacaccc atttgtacc 2040
catccaaagc acttaccagtt tagcaaatat taagagagct ggtagctcat catctacca 2100
taaaagagaa aaaaattcgg aagtgatca gagcacttat tcaaatata gtatagaag 2160
ttcagaagag tcaccaaggt caaggagcag atcttctagg agtagatctt attccagatc 2220
atatcaaaag tcaactgagc tagctagtgc acattcaagg tctaggtctc catcatctag 2280
atctcattca cgaataaat acagtgaatc ttccagtggt agtagatcat ctctatatac 2340
ttctattagc agtcatgag gaaggcagc taagagagga cttagatcca gtggaaaaaa 2400
aaatagagct tcaataaaaa agcatagcag cagctctgaa aagacacttc acagtataa 2460
tgtcaaaagt agagacaggt ctctatgtgt gagaagatat agcagagaca gatcatctt 2520
agattattct tccagagagt agcagtcagc tggttcagcc acacagtcag ccaggaanaa 2580
agagagcagc ggcacaatgg aaagaacaca taataacaaa gaaaaaaca gaggtaaga 2640
aaaaatccag tctgaacggg aatgccctca ttcaaaaaaa agaacttga aagagaatc 2700
ttctgatcac cttagaaatg gcagttaagc caaaaggag aattatgctg gtagttaagt 2760
ggaactctgag tcaaatccag aacgagatgt cactaaaaaa agtaaaaatg actcccatcc 2820
atctctctgac aaggaagaag gtgagggcac atccgattct gaatcagagg ttagtgaat 2880
tcaaatcaaa gtcaaaccca caaccaagtc gtcccaaatc acttcactgc ctgatgata 2940
tgggtgttgg aaatcaagca aacagcgcac atcaacttct gactctgagg ggtcctgttc 3000
caattcggaa aacataggg gaagccaca aaagcacaaa catgggtcaa aggaataat 3060
taaaagagaa cacacccaaa agtgaaaga gaattgaaa gggaaaaaag acnaaaagca 3120
taaggctcca aaacgaagc aagcatttca ctggcagcct ccactagaat ttggtgaaga 3180
ggaggaggag gagattgatg acagcagat tactcaggaa tcaaaagaga aaaaagtctc 3240
tgaanaacaa gaacccataa agataatata tctaaaactc gagaatcca gtgaagagga 3300
cctttcaggt aaacatgata cagtgaactg tctatcagat cttagatcagt ttaactaaga 3360
tgatagtaaa ctcatgattt ctccacagc tttaataact gaggaaaatg tggcctgttt 3420
acnaaacatt cagcagtttg aagaaagtgt tcccaatgga gtggaagatg tgcttcaaac 3480
agatgaacac atgagagatc gcaactctga taggagttcc ccagcaaaag tagaggagac 3540
ttccctctca ggaatgcac ggccttgatc ccagatata aacattgttt tgaagcagga 3600
tatggcaacg gaacatcttc aagcagaggt agtaaaacag gaagcagca tgcocgaag 3660
taagtgttgg ggtgaagtgg ggaacacgga cagcagctct gctagcttgg ctagtgtcgg 3720
agaagagtcg ggggaagagg aggtgtgtga gaagagcagc atcaacttca ttgataagaa 3780
atggaagccc ctgcaaggtg tggggaacct ggacgcacct aatgctgcca catccagtgc 3840
tgtggaagtt aaggtgttga ccaactgtgc tgaatgaaa ccaaaagpct tgagaataga 3900
aatataaagc aaaaataaag ttccgacctg gtctctcttt gatgaagtaa gaagacagc 3960
acgcttaaac cgtagaccaa gaactcagga gagttaaggt gatgagcaga cgcctagtgc 4020
ggatgatgat agccagtcga ggaagtcag tagatctcga agtaaatctg aaaccaaatc 4080
aagacacagc acaaggtctg tctcctatag tcaactcaaga agtcgatoga gaagttccac 4140
atcatcttat cgaatcaaga gctactctag aagtcggagc agaggatggt acagcagagg 4200

```

WO 02/097032

PCT/US02/11152

```

ccgaaccaga agccggagca gttcctaccg gagttacaaa agtcacagga cgtccagcag 4260
gagcagatcc agggagcagct catatgatcc ccacagtcga tccagcaggt cctacaccta 4320
cgtatagctac tatagcagga gtcggagtcg aagttgaagc cagagaagtg acagttacca 4380
cggaggcaga agttataatc gccggtcocag gagttgtaga tcttatggct ctgacagtga 4440
aagtgaccga agttactctc atcacccggag cccacgtgag agccagcagat acagttgaaa 4500
acgtccggat acaaatlata tcttatttgt aaatatctgg caacttagct taagaaatgt 4560
aatgacagtc gtttcttcta tttaaatata agaggtgaat ttcaaaaata gacacttctt 4620
aattgttact gtttcttcta catgtggggg gaagaattta aaatacagat atgtctctca 4680
aaatatcttt tatgccatct ttacacagtg ccaactatgg aaatgaattt cattttcttg 4740
aatcaagaaa tegtgaattt tatctatgta taatttgcaa tattatttta agtctatttc 4800
actctatctt acgtatccct tagaatacag attctttttg cctgtttttc cagtttttag 4860
atatatgctg ccaagcatag aactgtgaag gagaactgtt aaagggggcc aaatatctat 4920
atactgatta catagagctt tgtacatatg tgccttaaaa acaaacccac cagaattgat 4980
actgttggtt accaggagta taaggcagtg gctctgggtt tcttaattca ttcttaacct 5040
cttgatactt tccacaggat aggaagtggt tcatcataca tccacacacg tctgtattac 5100
ttcaggtctg tgggcaaggt taggaagaat caatcagcct taactataaa taactgcact 5160
gtctctgggg acttactatt ttatgttctt tttaactaat accgatcaga agtttaggtt 5220
ataaaaacca ttctacttca tgccttggtg cttggtaatt tttggtcgtt ctttaagcat 5280
tctcttata tatcatatat taaaatacca taaaaatgaa ttccagacaa aatcactggc 5340
accaaaaaag gtttattctg agctgtcttc actttgacta tttggggggc ttctctcaag 5400
taccagatgg ggttgggggc ccttgagcca ggccaggatt gcagtaagag atattggcca 5460
ctcaagctta ctgtgtgtgt gtgctctgg agagtggaag antggacttc aaagtaaca 5520
tcaaaaatct aactgccacc atcctggaga cattttgcag ggctttcttt tcaagtcttt 5580
caagtccagc atattaccac aacagcagct gaactgttgt aaccagcatg tttttctctat 5640
ttccactgtg acctgagctt gactcaagc cttgcgtgac ctgacccagg tgcagagac 5700
aggggaagag ggtatagagg tatagcataa attacatatt ttcatggctt tgggtgtgtc 5760
ctcaaaaaat aattggacct gtaaaaacta gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt 5820
gtgtgtgttt tttttttaat ctltactttg aatttgtttc ccaagtgtac ttaactacot 5880
tagtgcagct ttaatccagt tatgcagaag aaattcatat tgggtgcctg atgtagagct 5940
cagccaccac ctaccacagg ccttgtctgg tgtattttgg aagtggaaaa gagccctcag 6000
ttggaggagg ctgacaaccc ttggtgagag gaggtgtccc ttgaaatgat taaaactatc 6060
acccaagaaa ggtatgaaaa cagggtaaag ttgtcagttg ttbgcaggt caatagacag 6120
aaagtacatt agaaaacagg acttaggcct actggatact gaatacaaaa 6180
cagtattgat tatattaaag gtttccaaag gttgcctgca aaggagataa ttactactag 6240
tcagcaggaa aaaaatgcat tcagaaccca agcagaaaact gccaaatgta attaggttaa 6300
gaaaagtac ccttgggcag tgtattagtt ttctatttgt gtgtgacaaa ttaccccaaa 6360
tttagcagtt taaaaaaca taccatagc agttctgtag ctcatgagtc tggccacagt 6420
tggctggatt ctctgctcag ggttattaaag gctgaataaa ggggtggcag gacaacatct 6480
cttcatggag gctctgggga agaattctgt tctaaagtta ttacagttgt tggcggaatt 6540
cagttctttg ctguctctca gctggaggcc cctctctcac ctcaaggctg cctgcattcc 6600
ttcttatgtg gtccctctca gcttcaaaac agccttctct ctctttctca tgcctcatat 6660
ctctctcccg tctctctgtt ttagggcact atgattagct caagcccaca gatatttttt 6720
aaggttgatt gtgcgntaga acataattgc aggtgactgt tctcatctca tcatattcac 6780
gggttctgga gattagctca ttgaaagtgg gaggggcatl ttcaaatctt gctaccaca 6840
ggcaataact gcccatctca gctgtaggtg gaatttttac ccagaaaaga taggccttag 6900
aagcctcatt tcttttctcc atggaaaagg acagccctct gctgcagcgt tcaacttctg 6960
tgtttactga cagagtgaac tacagaaata gcttttcttc ctaaaagggga ttgtcttaca 7020
ttttgaagtt attttttaat aaattgaaat tatgttgt 7058

```

<210> 39
 <211> 1380
 <212> DNA
 <213> Homo sapiens

 <220>
 <221> misc_feature

WO 02/097032

PCT/US02/11152

<223> Incyte ID No: 7495168CB1

```

<400> 39
agggccggtc ttgcagagta gctgcgggtg gtgggctgtt ggcgcgagcg gtctggccca 60
agggctgggg gccggccgag ggtcttcggg agcaggcccg agggcgcgga gagatctctg 120
gatcccgctc cgcgcctgct aoccgccatg tcggcgagag cctcgggccc ggcggccgcc 180
ggcgcccgct cctggagagc ccccaagccc tcgggtctcg agcctggccc cgcgcctac 240
ggtctcaagc cgtgcacccc gaacagcaaa taagtgagag tgaactggg cggctctgtg 300
cactacacca cgtgcgcac cctcaaggga caggacacca tgcctcaagc catgttcagc 360
ggcgccgctc aggtgctgac cgtatccgga ggttgggtgc tgattgacgc gagcgccgct 420
cacttttgta caatctcaaa ttacctgggg gatgggtctg tgccactgoc ggagagtacg 480
agagaaactg gggagctgct gggcgaaaga cgtactacc tggtcaggg cctgattgag 540
gactgcccgc tggcgctgca gcaaaaaagg gagcgtgtt ccccgctgtg cctcatcccc 600
atggtgacat ctccccgga ggagcagcag ctctcgcca gccctccaa gcccgctgtg 660
aagctcctgc acaaccgcag taacaacaag tactctaca ccagcacttc agatgacaac 720
ctacttaaga acatcgagct gttcgacaag ctggccctgc gcttccacgg gcggctactc 780
ttctccaagg atgtcctggg ggaagagatc tgggtctgtt ctttctacgg gcaggccgcg 840
aaatcgccgc aggtgtgctg cactccatt gctctatgta cggagaagaa gcagaccagc 900
gtggaatttc cagaggcccg gatcttcgag gagaccctga acatctcat ctacgagact 960
ccccggggcc cagaccagc cctcctggag gccacgggg gagcagctgg agctgggtgg 1020
cctggccgag ggaagagatg agagaaccca gagaccgtg tcgcagagat caatgtccgg 1080
cgcatataca cccagcagca ggtctcctat ggcacaacaa ttgtcttcaa ggaactgact 1140
cctgaccttc cctgccttc ctcttgcctt ggaaccagat cctctctct ttcctcccc 1200
ttccccagat ttggcccggt ctctgtctgc caagtctgtg gtctctctct gtccctctat 1260
tgcatggcac agtcaactg gccctctct acccatccca acccatgct aacaaactg 1320
taactctgoc caccacttt cagccagcat atacatcttg tttctgctt gttctctgtc 1380

```

<210> 40

<211> 1773

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7483131CB1

```

<400> 40
ggccccacgg gccggctcag cggcggtggc ggcaggctgt ttttcttcaa ataaagaaca 60
tggtgaactt gatccacaca ttactgtatc atgggtgncg tgcacactgc tgtcctctct 120
ccttttccct cttggctact tgcctcttgg acaaaacaat tcgctgttac tctttacgtg 180
actttactga actgccacat tctccattga agtttcaata ctatgctgtc cactgctgct 240
gtttctcccc ttcaggacat attttggcat cgtgttcaac agatggtaac actgtctcat 300
ggaactactg aaatggcagc atgtgtggag tgatggnaac gccatgtggt agcctgttga 360
gggttttgca gttttcccca gactccacgt gtttggcaco aggggcagct gatggaactg 420
tggtttttgt gaatgcacag toatcccaat tatatagatg tggtagtggt aaagatggct 480
cctggcgccg atgtgcattt tctcctaagc gaagcttctt tgcactgggc tctcatgtgt 540
gtgatttaac agtgtgggat gataaaatga ggtgtctgca tagtgaaaaa gcacatgatc 600
ttggaattac ctgctgcgat ttttcttcc acocagtttc tgatggagaa caaggtcttc 660
agtttttttc actggcatca tgtggtcagg attgccaagt caaaatttgg attgtttctt 720
ttaccocatat cttaggttttt gaattaaaaa ataaaagtac actgagtggt cactgtgtct 780
ctgttctcgc ttgtgctttt tcccatgatg ggcagatgct agtctcaggg tcatgtggata 840
agtctgtcat agtatatgat actaatctgt agaatatact tcacacattg actcagcaca 900
ccaggatagt cacaacttgt gottttggca ctaataacct tttacttgtc actggttcaa 960
tggacaaaac agtgaacatc tggcaatttg acctggaac actttgccaa gcaaggagca 1020
cagaacatca gctgaagcaa tttaccgaag atcggtcaga ggaggatgtc tcaacatggc 1080
tttgtgcaca agatttaaaa gatctgtgtg gtattttcaa gatgaataac attgatggaa 1140

```

WO 02/097032

PCT/US02/11152

```

aagaactggt gaattttaca aaagaaaagt tggctgatga ttgaaaaatt gaattcttag 1200
gaactgctag taaagtctct aggaanaattg aagagctcag gaccaagggt aaatcccttt 1260
cttcagggaat tccctgatgaa ttatatatgt caataactag agaacttatg aaagatccgg 1320
tcactgcctc agatggctat tcaatatgaa aggaagcaat ggaatattgg atcagcaaaa 1380
agaaacgtac aagtcoccatg acaaatcttg ttcttccttc agcggtaatt acacaaataa 1440
ggactctgaa aatggccatc aatagatggc tggagacaca ccaaaagtaa aatgtttgat 1500
attgtattat ttatattttc agtgaatcga ttgaatgat ttataggtaa ataactaaca 1560
gcactattta aaagcaaaac aggaanaaggt taaactcttc aaatttagtt acctataaaa 1620
attgtcaatt ttcatctctt aaaaacacat ggaacttaata taaaagcctt ttgtactag 1680
tgaanaagat cttcagctat atagaataaa agttataact taaattgcaa aaaaaaaa 1740
aaaaaaatt ctggccgca atgaattcgt ggc 1773

```

```

<210> 41
<211> 2810
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 4558650CB1

```

```

<400> 41
actgccact tatgactga gccttcggga gggacagctc tacaggggtc ggcgctgta 60
aagacgtcac ttccgtcgga aggtttctcat gaggagccac tcagattcct gtcaattctc 120
aaagcattct cgtcgtgaaac atggccctgc caccattctt cggccagggt cggccaggcc 180
caacgcccco gacgcgcgcg cctcctgctc ctttcggctg tccgcacccg ccgtctgccc 240
cccgcgcttt ccgcgcgcct ctccccagc ggcgcggccc ttctccgggg gcctccgccc 300
ccttccttca gcctcgcctg gctctgcagc ccgagccctc cggcgaggcc tcccgcgccg 360
gaggcgggcc tggcgccctc taccgggtgc caccacgcgc cgtgcctcct ccgcgcgcgc 420
agtgctcgcc cttcccgggg accgacgcgc gcgagcgccc cgggcacccc cctcccgccc 480
cgggcgccgc ctggagcccg cggctggcctg agggcgccgc cgcgcggccc gacgtgctcg 540
gggagcgccc cctccacgc ctgcgcgcgc gcagtgagct ggagcgagtg ttccggagcc 600
cgcgcggggc aggtctgtcc gttccccagc gcacgcctgc cgggcccagc cttggcgag 660
tgcgcgcgcg attgtccgg gctctgcgcg tggatcgccg gctgcgcggc ctgagccagg 720
cctgcgcgga ggcggaagcc gacggcgccg cctgggtcct gctgtactcc cagacccgcc 780
cgctgcgcgc ggaactggcc gagcggtac agccgttgac ccaggctgac tatgtggcgc 840
agggcgggag gaggctggag aggttcgggc ggcgcggctc ggggtctcgc gagagggccc 900
gggaacgcga ggcgagcgag gaggcgagag ccgcgcgggc agtggaacgc gagcaggaga 960
ttgacgcgtg gagggtgaag tgtgtgcagg agtgaggaga gaagaagcgg gaggcaggaa 1020
tcaagcagc cgctgatggc gtactatctg aagttaggaa aaaaacagca gatccaaaa 1080
gaatggttga cattctacgg gctttggaga aattgaggaa actgaggaaa gaggctgcag 1140
cgaggaaagg ggtctgtcct ccagcctcag cagatgagac ttttacgcac catcttcagg 1200
gactgagaaa actcattaaa aagcgcctcg aactgtatga agctgaagag agagccctca 1260
gagtttatgct agagagagaa caagaggaag agaggaagaa agaattagaa aagaacaaa 1320
gaaaagagga agagaaaatt ttacttcaga aacgtgaat tgagttcaag ttgtttgggg 1380
atccagatga gttccacttt gctcacctct tggagccttt ccgacagtat tatctccaa 1440
ccgagcactc cctgccagcg ctcatccaga tcaggcatga ttgggataag tacctgggtc 1500
catccgatca tcccaaggcc aacttcgttc ccaaggatg ggtccttccc ccgtcccca 1560
gcaagcactc ctggccaact gctgttaagc tgcattagta aagatgctcc aggagtgctg 1620
tccagccagc gctctttcca cctgttaata ttacgatgg tgccatcttt gctgttagac 1680
taaatgcgaa cttctaaatt ccagtgtgga ttccctacc ctgaagttat gcttctctc 1740
tgtctctgt cctgcgcaaga ggtgcctctt gaatcagatt aatgtgggtt tccaggaaag 1800
gacttaggtg aactgaggtt tttaaccagc gaagtgaat acctgggttc accaaatttg 1860
cctctgtttt gaggggcttg gtccagagtg acctgttaat ttactctaac ttctttrgt 1920
gttgatgggt aagtacactc aaacactgaa tacaggtgtg tgatgggtag atttccagc 1980
cctctacta atagttagtg tgaagcagc cttgatgcaa aacctcctga ctttctctac 2040

```

WO 02/097032

PCT/US02/11152

```

ctgaagagcc ctttgacttc taggaagaaa ggtcaaaaat gttatottoa gttgtgttaa 2100
tcccagtttt agtgcagctt agggagctgc tagtttagaa gatggcagtg gctgtaggct 2160
gggtgtccag aaaaagatgg ggctagtctc tattattcag atggagaact tagaaaaact 2220
gaagagtacc caaatgggat tgtattttta tggcaaatgg ctgtattttt tccatgttag 2280
aaggatccta atgaagacac ctgtttatttt taagtttcta agggctotagt tgttcagaat 2340
ccccagggat atttccctaa cctcaactcag tccattgtta ggagccagtg tagctatgga 2400
attatcttag gaactcaagc ttctaaaaac atccattgta tcaaatctag gggaaaaagc 2460
aaataaaaa agtaaaattt ggccgggcac agtggctcac gctgttaac ccaacacttt 2520
gggagggcga ggggggcga tcaagaggtc aggaagtcac ggcactcctg gctaacacgg 2580
tgaacccttg tctctactaa aaatacaaaa aaatatatag tgggctagg tgggtgcac 2640
ctgtagtcac agctactctg ggggctgagg caggagaatg gtgtaaaaac caggagggag 2700
agcttgagat gagccagat ccgccacgg cactccagcc tgggagacag agcaagaact 2760
cgtctcaaaa aaaaaaagg taaaatttat tttttatatt cattaataaa 2810

```

<210> 42
 <211> 2549
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7506195CB1

```

<400> 42
aggttgaaag aetgcccact tatgcaactga gccctcggga gggacagtc tacaggggtc 60
gcgcgtctga agacgtcac ttccgtcggga aggttctcat gagagccac tcagattcct 120
gtcaattctc aaagcatctc cgtcgtgaac atggccctgc caccattctt cggccagggt 180
cgccacggcc caccgcccc gcagccggcc cctcctgctc ctttggctgt tccgccacgg 240
cgctgcctct ccccggtttt ccgcgcgcct ctccccagc ggcctggccc ttttcgggg 300
gcccgcgcct ccttccttaa gccctcgcgt gctctgcagc cccagagctc cggcgaggcc 360
tcccggggcg gaggggggcg tggcgcttcc taccgggtgc caccacggcc gctgcctcct 420
ccgcgcgcct agtctgggcc cttccgggg accgaagccg gcgaggggcc gcggccacgg 480
cctcccgccc cggggccgccc ctggagcccg cgggtggccc agggccgccc gccgcggccc 540
gacgtgctcg gggatgcggc cctccaaagc ctgcgcgacc ggcagtgact gggagcggtg 600
ttcgggaacc cggggcgggc aggcctgtcc gtgcccagc gcaagcatgc cgggccagc 660
cttggcgaa gtcgcgcgcg attgctccgg gctctgcgcc tgggtcgggc gctgcgcggc 720
ctgagccagg cctgcgcgga ggcgaagcc gacggcggg cctgggtcct gctgtactcc 780
cagaccggcg cgtgcgcgca ggaactggcc gagcggtcac agccgttgac ccaggttgc 840
tatgtggggc aggcgcggag gaggctggag agggctcggc gccgcgggct ggggcttgc 900
gagagggccc gggaaacgga ggcgcgcgg gaggcagagg ccgcgcgggc agtggaagc 960
gagcaggaga ttgaccgctg gagggtgaag tgtgtgcagg aggtggagga gaagaaggg 1020
gagcaggaa ccaagcagc cgtgatggc gtactatctg aagtgaggaa aaaaacagca 1080
gataccaaaa gaattgtgga cattctacgg gctttggaga aattgaggaa actgaggaaa 1140
gaggctgcag cgaggaaaga tgagttccca cttgctcacc tcttgaggcc ttccgacag 1200
tattatctcc aagccagaca ctcctgcaca ggcctcatcc agatcaggca tgattgggat 1260
cagtaacctg tgcctacgca tcaatccaaa ggcactctcg ttcccaagg atgggtcctt 1320
ccccgcctcc cagcaacgca catctgggca actgctgtta agctgcatta gtaagatgc 1380
tccaggagtg tgggtccagc agcgtctctt ccagctgtaa atattagcga tgggtgcact 1440
tttgtgtgta gactaaactc caacttctaa attccatgtg ctttccctcc acctgaagt 1500
tatgtcttcc ttctgtgctc tgtgttgccc agagggtgct cttgaatcag attaatgtg 1560
tttttcagga aaggaactag gtgaactgag gtttttccca caggacatga atgacctgg 1620
ttcaccaaat ttgcctctgt ttgaaggggc ttgggtccaga gtgaactgtt aatttactct 1680
aacttccctg tgtgttgatg ggtaaagtaa ctcaaacact gaatacaggt gtgtgatgg 1740
tagatttcc agccttctca ctcaatagta gtgtgaaggc aagcttgatg caaacctccc 1800
tgacctttcc taccgtgaaga gccctttgac ttctaggaag aaaggtcaaa aatgttatct 1860
tcagtgtgtt taatccagat tttagtgcag cttagggagg tgcagtttag gaagatggca 1920

```

WO 02/097032

PCT/US02/11152

```
gtggctgtag gctgggttgc cagaaaagat ggtggcctag ttttattatt cagatggaga 1980
acttagaaaa cctgaagagt acccaattg gattgtattt taatggacaa tggctgtatt 2040
ttttccatgt tagaaggatc ctaatgaag cactgttat ttttaagttt ctaagggtcc 2100
agtgtttcag aatccccaag gatatctccc taacctcact cagtcacatt gtaggggcca 2160
gtgtagctat ggaattatct taggaactca agcttctaaa actatccatg tagtcaatc 2220
taggggaaaa agcaaatata aatagtataa ttggccggg cactgtggct cagcctgta 2280
atcccaacac ttggggaggg cgaggcgggc cgtcacgag gtcaggagat caaggccatc 2340
ctggctaaca cggtagaacc ctgtctctac taaaaataca aaaaaatatt agctggggct 2400
aggtgggtga cactgtagt cccagctact ggggaggtg aggcaggaga atggtgtaaa 2460
accagaggag cagagcttgc agtgagccga gatggcgca oggcactca gcctgggaga 2520
cagagcaaga ctccgtctca aaaaaaaaaa 2549
```

【国際公開パンフレット（コレクトバージョン）】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number
WO 02/097032 A3(51) International Patent Classification: C07K 1/00,
A61K 38/00, 38/43, A23J 1/00, C12P 21/06, 21/04

(21) International Application Number: PCT/US02/11152

(22) International Filing Date: 5 April 2002 (05.04.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/282,110 6 April 2001 (06.04.2001) US
60/283,294 11 April 2001 (11.04.2001) US
60/286,820 26 April 2001 (26.04.2001) US
60/287,228 27 April 2001 (27.04.2001) US
60/291,662 16 May 2001 (16.05.2001) US
60/291,846 18 May 2001 (18.05.2001) US
60/293,727 25 May 2001 (25.05.2001) US
60/295,340 1 June 2001 (01.06.2001) US
60/295,263 1 June 2001 (01.06.2001) US
60/349,705 15 January 2002 (15.01.2002) US(71) Applicant (for all designated States except US): INCYTE
GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo
Alto, CA 94304 (US).

(72) Inventors and

(75) Inventors/Applicants (for US only): AZIMZAI, Valda
[US/US]; 5518 Boulder Canyon Drive, Castro Valley,
CA 94552 (US). AU-YOUNG, Janice, K. [US/US]; 233
Golden Eagle Lane, Brisbane, CA 94005 (US). BATRA,
Sajeev [US/US]; 555 El Camino Real #709, Sunny-
vale, CA 94087 (US). BAUGHN, Mariah, R. [US/US];
14244 Santiago Road, San Leandro, CA 94577 (US).
BECHA, Shanya, D. [US/US]; 21062 Gary Drive #117,
Castro Valley, CA 94546 (US). BOROWSKY, Mark, L.
[US/US]; 122 Orchard Avenue, Redwood City, CA 94061
(US). BURFORD, Neil [GB/US]; 105 Wildwood Circle,
Durham, CT 06422 (US). DING, Li [CN/US]; 3353
Alma Street #146, Palo Alto, CA 94306 (US). ELLIOTT,
Vicki, S. [US/US]; 3770 Polton Place Way, San Jose,
CA 95121 (US). EMERLING, Brooke, M. [US/US];
1735 Woodland Avenue #71, Palo Alto, CA 94303 (US).
GANDHI, Ameena, R. [US/US]; 705 5th Avenue, San
Francisco, CA 94118 (US). GIETZEN, Kimberly, J.
[US/US]; 691 Los Iluecos Drive, San Jose, CA 95123(US). GRIFFIN, Jennifer, A. [US/US]; 33691 Mello
Way, Fremont, CA 94555 (US). HAFALIA, April, J.A.
[US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054
(US). HONCHELL, Cynthia, D. [US/US]; 400 Laurel
Street #203, San Carlos, CA 94070 (US). LAL, Preeti, G.
[US/US]; P.O. Box 5142, Santa Clara, CA 95056 (US).
LEE, Soo, Yuen [KR/US]; 40 Westdale Avenue, Daly City,
CA 94015 (US). LU, Dyung, Aina, M. [US/US]; 233 Coy
Drive, San Jose, CA 95123 (US). ARVIZU, Chandra,
S. [US/US]; 1706 Morocco Drive, San Jose, CA 95125
(US). RAMKUMAR, Jayalaxmi [IN/US]; 34359 May-
bird Circle, Fremont, CA 94555 (US). REDDY, Roopa
[IN/US]; 1233 West McKinley Avenue #3, Sunnyvale,
CA 94086 (US). SANJANWALA, Madhu, M. [US/US];
210 Silvia Court, Los Altos, CA 94024 (US). TANG,
Y., Tom [US/US]; 4230 Ramwick Court, San Jose, CA
95118 (US). WALIA, Narinder, K. [US/US]; 890 Davis
Street A #205, San Leandro, CA 94577 (US). WANG,
Yu-mei, E. [CN/US]; 4624 Strawberry Park Drive, San
Jose, CA 95129 (US). WARREN, Bridget, A. [US/US];
2250 Homestead Court #2, Los Altos, CA 94024 (US).
XU, Yumeng [US/US]; 1739 Walnut Drive, Mountain
View, CA 94040 (US). YANG, Junming [CN/US]; 7125
Bark Lane, San Jose, CA 95129 (US). YAO, Monique, G.
[US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US).
YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA
94087 (US). ZEBARJADIAN, Yeganeh [IR/US]; 830
Junipero Serra Boulevard, San Francisco, CA 94127 (US).(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics,
Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).(81) Designated States (national): AI, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GI,
GM, GR, GU, HD, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MY, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VN, YU, ZA, ZM, ZW.(84) Designated States (regional): ARIPO patent (GI, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,
NE, SN, TD, TG).

[Continued on next page]

(54) Title: PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH

(57) Abstract: The invention provides human proteins associated with cell growth, differentiation, and death (CGDD) and polynu-
cleotides which identify and encode CGDD. The invention also provides expression vectors, host cells, antibodies, agonists, and
antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expres-
sion of CGDD.

WO 02/097032 A3

WO 02/097032 A3



Published:
— with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(88) Date of publication of the international search report:
15 May 2003

【国際調査報告】

INTERNATIONAL SEARCH REPORT		International application No. PCT/US02/11152												
A. CLASSIFICATION OF SUBJECT MATTER IPC(7) : C07K 1/00; A61K 38/00, 38/43; A22J 1/00; C12P 21/06, 21/04; US CL : 530/350+, 300, 412+; 435/69.1, 70.1; 514/12; 424/94.1+; According to International Patent Classification (IPC) or to both national classification and IPC														
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) U.S. : 530/350+, 300, 412+; 435/69.1, 70.1; 514/12; 424/94.1+; Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EAST, STN (MEDLINE, CAPLUS, EMBASE, BIOSIS)														
C. DOCUMENTS CONSIDERED TO BE RELEVANT <table border="1"> <thead> <tr> <th>Category *</th> <th>Citation of document, with indication, where appropriate, of the relevant passages</th> <th>Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>US 5,861,312 (VARSHAVSKY, et al) 19 January 1999 (19.01.1999), entire reference, especially column 5, lines 63-67 and column 10, lines 59-67.</td> <td>1, 2, 17, 18, and 56</td> </tr> <tr> <td>Y</td> <td>US 5,384,255 (CIECHANOVER, et al) 24 January 1995 (24.01.1995), entire reference, especially column 15, lines 61-68, column 16, lines 1-43, column 21, lines 34-49 and column 22, lines 10-26.</td> <td>19, 45</td> </tr> <tr> <td>A</td> <td>VARSHAVSKY, A. The N-end Rule: Functions, Mysteries, Uses. Proc. Natl. Acad. Sci. USA. October 1996, Vol 93, pages 12142-12149.</td> <td>19 and 45</td> </tr> </tbody> </table>			Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	X	US 5,861,312 (VARSHAVSKY, et al) 19 January 1999 (19.01.1999), entire reference, especially column 5, lines 63-67 and column 10, lines 59-67.	1, 2, 17, 18, and 56	Y	US 5,384,255 (CIECHANOVER, et al) 24 January 1995 (24.01.1995), entire reference, especially column 15, lines 61-68, column 16, lines 1-43, column 21, lines 34-49 and column 22, lines 10-26.	19, 45	A	VARSHAVSKY, A. The N-end Rule: Functions, Mysteries, Uses. Proc. Natl. Acad. Sci. USA. October 1996, Vol 93, pages 12142-12149.	19 and 45
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.												
X	US 5,861,312 (VARSHAVSKY, et al) 19 January 1999 (19.01.1999), entire reference, especially column 5, lines 63-67 and column 10, lines 59-67.	1, 2, 17, 18, and 56												
Y	US 5,384,255 (CIECHANOVER, et al) 24 January 1995 (24.01.1995), entire reference, especially column 15, lines 61-68, column 16, lines 1-43, column 21, lines 34-49 and column 22, lines 10-26.	19, 45												
A	VARSHAVSKY, A. The N-end Rule: Functions, Mysteries, Uses. Proc. Natl. Acad. Sci. USA. October 1996, Vol 93, pages 12142-12149.	19 and 45												
☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.														
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance. "B" earlier application or patent published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other cited documents, such combination being obvious to a person skilled in the art "Z" document member of the same patent family														
Date of the actual completion of the international search 12 December 2002 (12.12.2002)		Date of mailing of the international search report 15 JAN 2003												
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230		Authorized officer: Deborah J. Reynolds Laping Chen, Ph.D. Telephone No. 703-305-2785 SUPERVISORY PATENT EXAMINER TECHNOLOGY CENTER 1600												

INTERNATIONAL SEARCH REPORT		International application No. PCT/US02/11152
Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)		
This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:		
1.	☐	Claim Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
2.	☐	Claim Nos.: because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3.	☐	Claim Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)		
This International Searching Authority found multiple inventions in this international application, as follows: Please See Continuation Sheet		
1.	☐	As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2.	☐	As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3.	☐	As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4.	☒	No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1, 2, 17-19, 45 and 56 for SEQ ID:1
Remark on Protest		
	☐	The additional search fees were accompanied by the applicant's protest.
	☐	No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

PCT/US02/11152

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claim(s) 1, 2, 17-19, 45 and 56-76, drawn to a polypeptide comprising an amino acid sequence selected from any one of SEQ ID NO:1-21, a composition comprising a respective polypeptide, a method purifying a polypeptide, and a method for treating a disease associated with decreased expression of functional CGDD.

Group II, claim(s) 3-7, 9, 10, 12-15 and 77-97, drawn to isolated polynucleotide comprising polynucleotide sequence selected from any one of SEQ ID NO:22-42, a vector, a cell transformed with the respective selected polynucleotide, and a method of producing a polypeptide

Group III, claim(s) 8, drawn to a transgenic organism comprising a recombinant polynucleotide expressing polypeptide selected from any one of SEQ ID NO:1-21.

Group IV, claim(s) 11, 30-44, drawn to an isolated antibody which specifically binds to a polypeptide of any one SEQ ID NO:1-21, a composition comprising antibody, a diagnostic test for a conditions associated with the expression of CGDD in a biological sample using the antibody.

Group V, claim(s) 12 and 16, drawn to a method of detecting a target polynucleotide in a sample using polymerase chain reaction amplification.

Group VI, claim(s) 20, drawn to a method of screening a compound for effectiveness as an agonist of a polypeptide selected from any one of SEQ ID NO:1-21.

Group VII, claim(s) 21 and 22, drawn to a composition comprising an agonist compound identified by a method claim 20, and a method for treating a disease associated with decrease expression of functional CGDD using the composition.

Group VIII, claim(s) 23, drawn to a method of screening a compound for effectiveness as an antagonist of a polypeptide selected from any one of SEQ ID NO:1-21.

Group IX, claim(s) 24 and 25, drawn to a composition comprising an antagonist compound identified by a method claim 23, and a method for treating a disease associated with overexpression of functional CGDD using the composition.

Group X, claim(s) 26, drawn to a method of screening for a compound that specifically binds to the polypeptide selected from any one of SEQ ID NO:1-21.

Group XI, claim(s) 27, drawn to a method of screening for a compound that modulates the activity of the polypeptide selected from any one of SEQ ID NO:1-21.

Group XII, claim(s) 28, drawn to a method of screening a compound for effectiveness in altering expression of a target polynucleotide selected from any one of SEQ ID NO:22-42.

Group XIII, claim(s) 29, drawn to a method of assessing toxicity of a test compound by treating a biological sample containing nucleic acids with the test compound, and hybridizing the biological sample with at least 20 contiguous nucleotide of selected from any one of polynucleotides of claim 12.

INTERNATIONAL SEARCH REPORT

PCT/US02/11152

Group XIV, claim(s) 46 and 47, drawn to a microarray wherein at least one element of the microarray is a polynucleotide of claim 13, which comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12, and a method of generating an expression profile of a sample using the microarray.

Group XV, claim(s) 48-55, an array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of any one of the polynucleotide of claim 12.

Each Group contains claims directed to more than one invention distinguished by each sequence. These inventions are deemed to lack unity of invention because they are not so linked as to form a single general inventive concept under PCT Rule 13.1.

In order for more than one invention to be examined, the appropriate additional examination fees must be paid. The inventions distinguished by different sequences in each group are as follows:

Groups I, III, IV and VI-XI contain SEQ ID NO:1-21; 21 sequences: $9 \times 21 = 189$ inventions
Groups II, V and XII- XV contain SEQ ID NO:22-42; 21 sequences: $6 \times 21 = 126$ inventions

The inventions listed as Groups I-XV do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Group I encompasses a polypeptide comprising an amino acid sequence selected from any one of SEQ ID NO:1-21, a composition comprising a respective polypeptide, and the first method for treating a disease associated with decreased expression of functional CGDD. Groups II-XV are directed to different products or methods that require different special technical features as summarized as follows: Group III and XV, are directed to different products such as a transgenic organism, or an array, respectively; Groups II and IV-XIV are directed to different products and methods, each method require specific technique feature other than the technique used in Group I. Moreover, the nucleic acids with >90% or >95% identical to polynucleotide sequences of instant invention are well known in the art, an example of which is N_Geneseq access number AAF21924, described by Rosen et al. (WO 00/55173, published 21 September 2000). Further, different sequences are differ at chemical structure and associated function. Inventions containing different sequences are directed to functionally different inventions that require different technique feature to accomplish. Thus, Groups I-XV (inventions 1-315) lack a common special technical feature. Further, 37 CFR 1.475 does not provide for multiple independent products, methods of manufacture and methods of use (37 CFR 1.475(d)). Therefore, The inventions listed as Groups I-XV do not relate to a single general inventive concept under PCT Rule 13.1.

フロントページの続き

(51)Int.Cl.⁷

F I

テーマコード (参考)

A 6 1 P 3/10
 A 6 1 P 7/06
 A 6 1 P 9/00
 A 6 1 P 9/10
 A 6 1 P 11/00
 A 6 1 P 11/06
 A 6 1 P 15/00
 A 6 1 P 17/00
 A 6 1 P 19/00
 A 6 1 P 19/02
 A 6 1 P 19/06
 A 6 1 P 21/00
 A 6 1 P 21/04
 A 6 1 P 25/00
 A 6 1 P 25/08
 A 6 1 P 25/16
 A 6 1 P 25/18
 A 6 1 P 25/20
 A 6 1 P 25/28
 A 6 1 P 27/06
 A 6 1 P 27/12
 A 6 1 P 29/00
 A 6 1 P 35/00
 A 6 1 P 37/00
 A 6 1 P 37/08
 C 0 7 K 1/22
 C 0 7 K 14/47
 C 0 7 K 14/525
 C 0 7 K 16/18
 C 0 7 K 16/46
 C 1 2 M 1/00
 C 1 2 M 1/34
 C 1 2 N 1/15
 C 1 2 N 1/19
 C 1 2 N 1/21
 C 1 2 N 5/10
 C 1 2 N 9/00
 C 1 2 N 15/02
 C 1 2 P 21/02
 C 1 2 P 21/08
 C 1 2 Q 1/68
 G 0 1 N 33/15
 G 0 1 N 33/50
 G 0 1 N 33/53
 G 0 1 N 33/566

A 6 1 P 3/10
 A 6 1 P 7/06
 A 6 1 P 9/00
 A 6 1 P 9/10
 A 6 1 P 11/00
 A 6 1 P 11/06
 A 6 1 P 15/00
 A 6 1 P 17/00
 A 6 1 P 19/00
 A 6 1 P 19/02
 A 6 1 P 19/06
 A 6 1 P 21/00
 A 6 1 P 21/04
 A 6 1 P 25/00
 A 6 1 P 25/08
 A 6 1 P 25/16
 A 6 1 P 25/18
 A 6 1 P 25/20
 A 6 1 P 25/28
 A 6 1 P 27/06
 A 6 1 P 27/12
 A 6 1 P 29/00
 A 6 1 P 29/00
 A 6 1 P 35/00
 A 6 1 P 37/00
 A 6 1 P 37/08
 C 0 7 K 1/22
 C 0 7 K 14/47
 C 0 7 K 14/525
 C 0 7 K 16/18
 C 0 7 K 16/46
 C 1 2 M 1/00
 C 1 2 M 1/34
 C 1 2 N 1/15
 C 1 2 N 1/19
 C 1 2 N 1/21
 C 1 2 N 9/00
 C 1 2 P 21/02
 C 1 2 P 21/08
 C 1 2 Q 1/68
 G 0 1 N 33/15
 G 0 1 N 33/50
 G 0 1 N 33/53
 G 0 1 N 33/566
 C 1 2 N 15/00
 C 1 2 N 15/00

1 0 1

A
 Z

 C
 A
 Z
 Z
 D
 M

 C
 F

C 1 2 N 5/00

A

A 6 1 K 37/02

- (31)優先権主張番号 60/287,228
(32)優先日 平成13年4月27日(2001.4.27)
(33)優先権主張国 米国(US)
(31)優先権主張番号 60/291,662
(32)優先日 平成13年5月16日(2001.5.16)
(33)優先権主張国 米国(US)
(31)優先権主張番号 60/291,846
(32)優先日 平成13年5月18日(2001.5.18)
(33)優先権主張国 米国(US)
(31)優先権主張番号 60/293,727
(32)優先日 平成13年5月25日(2001.5.25)
(33)優先権主張国 米国(US)
(31)優先権主張番号 60/295,340
(32)優先日 平成13年6月1日(2001.6.1)
(33)優先権主張国 米国(US)
(31)優先権主張番号 60/295,263
(32)優先日 平成13年6月1日(2001.6.1)
(33)優先権主張国 米国(US)
(31)優先権主張番号 60/349,705
(32)優先日 平成14年1月15日(2002.1.15)
(33)優先権主張国 米国(US)

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT, BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN, TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EC,EE,ES,FI,GB,GD,GE, GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NO,NZ,OM,PH,P L,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZM,ZW

- (72)発明者 オウーヤング、ジャニス・ケイ
アメリカ合衆国カリフォルニア州94005・ブリスベン・ゴールドエンイーグルレーン 233
(72)発明者 バトラ、サジープ
アメリカ合衆国カリフォルニア州94087・サニーベイル・#709・エルカミーノリアル 55
(72)発明者 ボーゲン、マライア・アール
アメリカ合衆国カリフォルニア州94577・サンレアンドロ・サンティアゴロード 14244
(72)発明者 ベチャ、シャニア・ディー
アメリカ合衆国カリフォルニア州94546・カストロバレー・#117・ゲイリードライブ 21062
(72)発明者 ボロースキー、マーク・エル
アメリカ合衆国カリフォルニア州94061・レッドウッドシティー・オーチャードアベニュー 122
(72)発明者 バーフォード、ニール
アメリカ合衆国コネチカット州06422・ダラム・ワイルドウッドサークル 105
(72)発明者 ディング、リー
アメリカ合衆国カリフォルニア州94306・パロアルト・#146・アルマストリート 3353
(72)発明者 エリオット、ビッキー・エス

- アメリカ合衆国カリフォルニア州 9 5 1 2 1 ・サンノゼ・ポルトンプレイスウェイ 3 7 7 0
(72)発明者 エマーリング、ブルック・エム
アメリカ合衆国カリフォルニア州 9 4 3 0 3 ・パロアルト・# 7 1 ・ウッドランドアベニュー 1
7 3 5
- (72)発明者 ガンディー、アミーナ・アール
アメリカ合衆国カリフォルニア州 9 4 1 1 8 ・サンフランシスコ・フィフスアベニュー 7 0 5
- (72)発明者 ギエツェン、キンバリー・ジェイ
アメリカ合衆国カリフォルニア州 9 5 1 2 3 ・サンノゼ・ロスウエコスドライブ 6 9 1
- (72)発明者 グリフィン、ジェニファー・エイ
アメリカ合衆国カリフォルニア州 9 4 5 5 5 ・フレモント・メローウェイ 3 3 6 9 1
- (72)発明者 ハファリア、エープリル、ジェイ・エイ
アメリカ合衆国カリフォルニア州 9 5 0 5 4 ・サンタクララ・コーレデブリマベータ 2 2 2 7
- (72)発明者 ホンシェル、シンシア・ディー
アメリカ合衆国カリフォルニア州 9 4 0 7 0 ・サンカルロス・# 2 0 3 ・ローレルストリート 4
0 0
- (72)発明者 ラル、ブリーティ・ジー
アメリカ合衆国カリフォルニア州 9 5 0 5 6 ・サンタクララ・ピーオーボックス 5 1 4 2
- (72)発明者 リー、ソー・ユーン
アメリカ合衆国カリフォルニア州 9 4 0 1 5 ・デイリーシティ・ウエストデールアベニュー 4 0
- (72)発明者 リュ、デュング・アイナ・エム
アメリカ合衆国カリフォルニア州 9 5 1 2 3 ・サンノゼ・コイドライブ 2 3 3
- (72)発明者 アービズ、チャンドラ・エス
アメリカ合衆国カリフォルニア州 9 5 1 2 5 ・サンノゼ・モロッコドライブ 1 7 0 6
- (72)発明者 ランクマール、ジャヤラクシ
アメリカ合衆国カリフォルニア州 9 4 5 5 5 ・フレモント・メイバードサークル 3 4 3 5 9
- (72)発明者 レディ、ルーパ
アメリカ合衆国カリフォルニア州 9 4 0 8 6 ・サニーバイル・# 3 ・ウェストマッキンレーアベニ
ュー 1 2 3 3
- (72)発明者 サンジャンワラ、マデュー・エム
アメリカ合衆国カリフォルニア州 9 4 0 2 4 ・ロスアルトス・シルビアコート 2 1 0
- (72)発明者 タング、ワイ・トム
アメリカ合衆国カリフォルニア州 9 5 1 1 8 ・サンノゼ・ランウィックコート 4 2 3 0
- (72)発明者 チョーラ、ナリンダー・ケイ
アメリカ合衆国カリフォルニア州 9 4 5 8 7 ・ユニオンシティ・# 7 1 2 ・ユニオンスクエア 3
3
- (72)発明者 ワング、ユーメイ・イー
アメリカ合衆国カリフォルニア州 9 5 1 2 9 ・サンノゼ・ストロベリーパークドライブ 4 6 2 4
- (72)発明者 ワレン、ブリジット・エイ
アメリカ合衆国カリフォルニア州 9 4 0 2 4 ・ロスアルトス・# 2 ・ホームステッドコート 2 2
5 0
- (72)発明者 スー、ユーミング
アメリカ合衆国カリフォルニア州 9 4 0 4 0 ・マウンテンビュー・ウォルナットドライブ 1 7 3
9
- (72)発明者 ヤング、ジュンミング
アメリカ合衆国カリフォルニア州 9 5 1 2 9 ・サンノゼ・パークレーン 7 1 2 5
- (72)発明者 ヤオ、モニーク・ジー
アメリカ合衆国インディアナ州 4 6 0 3 3 ・カーメル・ウッドゲートドライブ・ 1 1 8 9
- (72)発明者 ユエ、ヘンリー
アメリカ合衆国カリフォルニア州 9 4 0 8 7 ・サニーバイル・ルイスアベニュー 8 2 6
- (72)発明者 ゼバージャディアン、イエガネー

アメリカ合衆国カリフォルニア州 9 4 1 2 7 ・ サンフランシスコ ・ ジュニペロセラブルバード
8 3 0

F ターム(参考) 2G045 AA35 AA40 BA11 BB50 DA13 DA36 FB02
4B024 AA11 AA19 BA07 BA28 BA43 BA80 CA04 CA05 CA06 CA09
DA02 EA02 EA04 FA02 FA10 GA05 GA11 GA18 GA19 HA03
HA08 HA12 HA14 HA19
4B029 AA07 AA21 AA23 BB20 CC03 FA12
4B050 CC03 DD11 LL03
4B063 QA01 QA18 QA19 QA20 QQ03 QQ20 QQ42 QR08 QR32 QR42
QR55 QR62 QR82 QS10 QS12 QS16 QS25 QS34 QS36 QX02
4B064 AG01 AG07 AG27 AG31 CA10 CA19 CA20 CC24 CE12 DA13
4B065 AA90X AA91Y AA93Y AB01 AC14 BA02 BA24 CA24 CA46
4C084 AA02 AA07 AA17 BA22 MA24 MA52 MA56 MA66 NA14 ZA011
ZA031 ZA051 ZA061 ZA151 ZA161 ZA331 ZA361 ZA451 ZA551 ZA591
ZA681 ZA811 ZA891 ZA941 ZA961 ZB071 ZB111 ZB131 ZB151 ZB261
ZB331 ZC351 ZC781
4C085 AA13 AA14 BB11 CC04 CC05 DD22 DD23 DD33 DD43 EE01
GG01
4H045 AA10 AA11 AA20 AA30 AA40 BA10 CA40 DA76 DA86 EA50
FA72 FA74 GA26

专利名称(译)	细胞生长，分化和细胞死亡相关的蛋白质		
公开(公告)号	JP2004533829A	公开(公告)日	2004-11-11
申请号	JP2003500201	申请日	2002-04-05
[标]申请(专利权)人(译)	洞察Genomics公司		
申请(专利权)人(译)	洞察基因组公司		
[标]发明人	アジムザイヤルダ オウヤングジャンスケイ バトラサジープ ボーグンマライアール ベチャシャニアディー ボロースキーマークエル バーフォードニール ディングリー エリオットビッキーエス エマーリングブルックエム ガンディーアミーナール ギエツエンキンバリージェイ グリフィンジェニファーエイ ハファリアエープリルジェイエイ ホンシェルシンシアディー ラルプリーティジー リーソーユーン リュデュングアイナエム アービズチャンドラエス ランクマールジャヤラクシミ レディルーバ サンジャンワラマデューエム タングワイトム チョーラナリンダーケイ ワングユーメイイー ワレンブリジットエイ スーユーミング ヤングジュンミング ヤオモニークジー ユエヘンリー ゼバージャディアンイエガネー		
发明人	アジムザイ、ヤルダ オウ-ヤング、ジャンス・ケイ バトラ、サジープ ボーグン、マライア・アール ベチャ、シャニア・ディー ボロースキー、マーク・エル バーフォード、ニール ディング、リー エリオット、ビッキー・エス エマーリング、ブルック・エム ガンディー、アミーナ・アール ギエツエン、キンバリー・ジェイ グリフィン、ジェニファー・エイ		

ハファリア、エープリル、ジェイ・エイ
ホンシエル、シンシア・ディー
ラル、プリーティ・ジー
リー、ソー・ユーン
リュ、デュング・アイナ・エム
アービズ、チャンドラ・エス
ランクマール、ジャヤラクシミ
レディ、ルーパ
サンジャンワラ、マデュー・エム
タング、ワイ・トム
チョーラ、ナリンダー・ケイ
ワング、ユーメイ・イー
ワレン、ブリジット・エイ
スー、ユーミング
ヤング、ジュンミング
ヤオ、モニー・ク・ジー
ユエ、ヘンリー
ゼバージャディアン、イエガネー

IPC分類号	G01N33/50 A61K38/00 A61K39/395 A61K45/00 A61P1/04 A61P3/10 A61P7/06 A61P9/00 A61P9/10 A61P11/00 A61P11/06 A61P15/00 A61P17/00 A61P19/00 A61P19/02 A61P19/06 A61P21/00 A61P21/04 A61P25/00 A61P25/08 A61P25/16 A61P25/18 A61P25/20 A61P25/28 A61P27/06 A61P27/12 A61P29/00 A61P35/00 A61P37/00 A61P37/08 C07K1/22 C07K14/47 C07K14/525 C07K16/18 C07K16/46 C12M1/00 C12M1/34 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N9/00 C12N15/02 C12N15/09 C12P21/02 C12P21/08 C12Q1/68 G01N33/15 G01N33/53 G01N33/566
CPC分類号	A61K38/00 A61P1/04 A61P3/10 A61P7/06 A61P9/00 A61P9/10 A61P11/00 A61P11/06 A61P15/00 A61P17/00 A61P19/00 A61P19/02 A61P19/06 A61P21/00 A61P21/04 A61P25/00 A61P25/08 A61P25/16 A61P25/18 A61P25/20 A61P25/28 A61P27/06 A61P27/12 A61P29/00 A61P35/00 A61P37/00 A61P37/08 C07K14/4738
FI分類号	C12N15/00.ZNA.A A61K39/395.D A61K39/395.N A61K45/00 A61P1/04 A61P3/10 A61P7/06 A61P9/00 A61P9/10 A61P11/00 A61P11/06 A61P15/00 A61P17/00 A61P19/00 A61P19/02 A61P19/06 A61P21/00 A61P21/04 A61P25/00 A61P25/08 A61P25/16 A61P25/18 A61P25/20 A61P25/28 A61P27/06 A61P27/12 A61P29/00 A61P29/00.101 A61P35/00 A61P37/00 A61P37/08 C07K1/22 C07K14/47 C07K14/525 C07K16/18 C07K16/46 C12M1/00.A C12M1/34.Z C12N1/15 C12N1/19 C12N1/21 C12N9/00 C12P21/02.C C12P21/08 C12Q1/68.A G01N33/15.Z G01N33/50.Z G01N33/53.D G01N33/53.M G01N33/566 C12N15/00.C C12N15/00.F C12N5/00.A A61K37/02
F-TERM分類号	2G045/AA35 2G045/AA40 2G045/BA11 2G045/BB50 2G045/DA13 2G045/DA36 2G045/FB02 4B024/AA11 4B024/AA19 4B024/BA07 4B024/BA28 4B024/BA43 4B024/BA80 4B024/CA04 4B024/CA05 4B024/CA06 4B024/CA09 4B024/DA02 4B024/EA02 4B024/EA04 4B024/FA02 4B024/FA10 4B024/GA05 4B024/GA11 4B024/GA18 4B024/GA19 4B024/HA03 4B024/HA08 4B024/HA12 4B024/HA14 4B024/HA19 4B029/AA07 4B029/AA21 4B029/AA23 4B029/BB20 4B029/CC03 4B029/FA12 4B050/CC03 4B050/DD11 4B050/LL03 4B063/QA01 4B063/QA18 4B063/QA19 4B063/QA20 4B063/QQ03 4B063/QQ20 4B063/QQ42 4B063/QR08 4B063/QR32 4B063/QR42 4B063/QR55 4B063/QR62 4B063/QR82 4B063/QS10 4B063/QS12 4B063/QS16 4B063/QS25 4B063/QS34 4B063/QS36 4B063/QX02 4B064/AG01 4B064/AG07 4B064/AG27 4B064/AG31 4B064/CA10 4B064/CA19 4B064/CA20 4B064/CC24 4B064/CE12 4B064/DA13 4B065/AA90X 4B065/AA91Y 4B065/AA93Y 4B065/AB01 4B065/AC14 4B065/BA02 4B065/BA24 4B065/CA24 4B065/CA46 4C084/AA02 4C084/AA07 4C084/AA17 4C084/BA22 4C084/MA24 4C084/MA52 4C084/MA56 4C084/MA66 4C084/NA14 4C084/ZA011 4C084/ZA031 4C084/ZA051 4C084/ZA061 4C084/ZA151 4C084/ZA161 4C084/ZA331 4C084/ZA361 4C084/ZA451 4C084/ZA551 4C084/ZA591 4C084/ZA681 4C084/ZA811 4C084/ZA891 4C084/ZA941 4C084/ZA961 4C084/ZB071 4C084/ZB111 4C084/ZB131 4C084/ZB151 4C084/ZB261 4C084/ZB331 4C084/ZC351 4C084/ZC781 4C085/AA13 4C085/AA14 4C085/BB11 4C085/CC04 4C085/CC05 4C085/DD22 4C085/DD23 4C085/DD33 4C085/DD43 4C085/EE01 4C085/GG01 4H045/AA10 4H045/AA11 4H045/AA20 4H045/AA30 4H045/AA40 4H045/BA10 4H045/CA40 4H045/DA76 4H045/DA86 4H045/EA50 4H045/FA72 4H045/FA74 4H045/GA26

優先権

60/282110 2001-04-06 US
60/283294 2001-04-11 US
60/286820 2001-04-26 US
60/287228 2001-04-27 US
60/291662 2001-05-16 US
60/291846 2001-05-18 US
60/293727 2001-05-25 US
60/295340 2001-06-01 US
60/295263 2001-06-01 US
60/349705 2002-01-15 US

外部リンク

Espacenet

摘要(译)

本発明提供了与细胞生长，分化和死亡（CGDD）有关的人蛋白质以及鉴定和编码CGDD的多核苷酸。 本發明还提供表达载体，宿主细胞，抗体，激动剂和拮抗剂。 本發明还提供了用于诊断，治疗或预防与CGDD异常表达有关的疾病的方法。

(P2004-533629A)			
(43) 公表日 平成16年11月11日(2004.11.11)			
(51) Int. Cl. ⁷	F 1	ターマコード (参考)	
C 1 2 N 15/09	C 1 2 N 15/00	Z N A A	2 G 0 4 5
A 6 1 K 38/00	A 6 1 K 39/395	D	4 B 0 2 4
A 6 1 K 39/395	A 6 1 K 39/395	N	4 B 0 2 9
A 6 1 K 45/00	A 6 1 K 45/00		4 B 0 5 0
A 6 1 P 1/04	A 6 1 P 1/04		4 B 0 6 3
審査請求 未請求 予備審査請求 未請求 (全 396 頁) 最終頁に続く			
(21) 出願番号	特願2003-500201 (P2003-500201)	(71) 出願人	301005050
(86) (22) 出願日	平成14年4月5日 (2002.4.5)		インサイト・ゲノミックス・インコーポレ
(85) 翻訳文提出日	平成15年10月2日 (2003.10.2)		イテッド
(86) 国際出願番号	PCT/US2002/011152		アメリカ合衆国カリフォルニア州9430
(87) 国際公開番号	W02002/097032		4・パロアルト・ボートドライブ 31
(87) 国際公開日	平成14年12月5日 (2002.12.5)		60
(31) 優先権主張番号	60/282,110	(74) 代理人	100089266
(32) 優先日	平成13年4月6日 (2001.4.6)		弁理士 大島 陽一
(33) 優先権主張国	米国 (US)		アジムザイ、ヤルダ
(31) 優先権主張番号	60/283,294	(72) 発明者	アメリカ合衆国カリフォルニア州9455
(32) 優先日	平成13年4月11日 (2001.4.11)		2・カストロバレー・ボールダーキャニオ
(33) 優先権主張国	米国 (US)		ンドライブ 5518
(31) 優先権主張番号	60/286,820		
(32) 優先日	平成13年4月26日 (2001.4.26)		
(33) 優先権主張国	米国 (US)		最終頁に続く