

(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2004-510407

(P2004-510407A)

(43) 公表日 平成16年4月8日(2004.4.8)

(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	39/00	Z	4 B O 2 4
A 6 1 K	38/43	A 6 1 K	45/00		4 B O 5 0
A 6 1 K	39/00	A 6 1 K	49/00	Z	4 B O 6 3
A 6 1 K	45/00	A 6 1 P	29/00		4 B O 6 4
		審査請求	未請求	予備審査請求	有 (全 190 頁) 最終頁に続く

(21) 出願番号	特願2001-587126 (P2001-587126)	(71) 出願人	301005050
(86) (22) 出願日	平成13年5月22日 (2001.5.22)		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14年11月22日 (2002.11.22)		イテッド
(86) 国際出願番号	PCT/US2001/016808		アメリカ合衆国カリフォルニア州9430
(87) 国際公開番号	W02001/090330		4・パロアルト・ポータードライブ 31
(87) 国際公開日	平成13年11月29日 (2001.11.29)		60
(31) 優先権主張番号	60/207, 248	(74) 代理人	100089266
(32) 優先日	平成12年5月25日 (2000.5.25)		弁理士 大島 陽一
(33) 優先権主張国	米国 (US)	(72) 発明者	ユエ、ヘンリー
(31) 優先権主張番号	60/208, 791		アメリカ合衆国カリフォルニア州9408
(32) 優先日	平成12年6月1日 (2000.6.1)		7・サニーバイル・ルイスアベニュー 8
(33) 優先権主張国	米国 (US)		26
(31) 優先権主張番号	60/210, 585		
(32) 優先日	平成12年6月8日 (2000.6.8)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 アミノアシル tRNA 合成酵素

(57) 【要約】

本発明はアミノアシル tRNA 合成酵素 (A T R S) および A T R S 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、A T R S 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 (d) を有する群から選択した実質上単離されたポリペプチド。

(a) SEQ ID NO: 1 - 4 (配列番号 1 乃至 4) を有する群から選択したアミノ酸配列を含むポリペプチド

(b) SEQ ID NO: 1 - 4 を有する群から選択したアミノ酸配列と少なくとも 90 % が同一であるようなアミノ酸配列を有する天然のポリペプチド

(c) SEQ ID NO: 1 - 4 を有す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

(d) SEQ ID NO: 1 - 4 を有す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抗原性断片 10

【請求項 2】

SEQ ID NO: 1 - 4 を有する群から選択した請求項 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: 5 - 8 (配列番号 5 乃至 8) を有する群から選択した請求項 4 に記載の単離されたポリヌクレオチド。 20

【請求項 6】

請求項 3 に記載のポリヌクレオチドに機能的に結合したプロモーター配列を含む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含む遺伝形質転換体。

【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、 30

(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ポリヌクレオチドに機能的に結合されたプロモーター配列を含む組換えポリヌクレオチドで形質転換された細胞を培養するステップと、

(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り、前記組換えポリヌクレオチドが、請求項 1 に記載の前記ポリペプチドをコードするポリヌクレオチドに機能的に結合したプロモーター配列を有することを特徴とする方法。

【請求項 10】

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 11】

以下の (a) 乃至 (d) を有する群から選択した単離されたポリヌクレオチド。 40

(a) SEQ ID NO: 5 - 8 を有する群から選択した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: 5 - 8 を有す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 % が同一であるような天然のポリヌクレオチド

(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

(e) (a) ~ (d) の RNA 等価物

【請求項 12】

請求項 11 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含む単離されたポリヌクレオチド。 50

【請求項 13】

請求項 11 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配列を有する少なくとも 20 の連続したヌクレオチドを含むプローブを用いて前記サンプルを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存在・不存在を検出し、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からなり、前記プローブと前記標的ポリヌクレオチドまたは断片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、前記プローブが前記標的ポリヌクレオチドに特異的にハイブリダイズすることを特徴とする方法。

10

【請求項 14】

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 13 に記載の方法。

【請求項 15】

請求項 11 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) 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) 前記標的ポリヌクレオチドまたはその断片の存在・不存在を検出し、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を含むことを特徴とする方法。

20

【請求項 16】

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有することを特徴とする成分。

【請求項 17】

前記ポリペプチドが、SEQ ID NO: 1 - 4 を有する群から選択したアミノ酸配列を含むことを特徴とする請求項 16 に記載の成分。

【請求項 18】

機能的な ATR S の発現の低下に関連する疾患や病態の治療方法であって、そのような治療が必要な患者に請求項 16 の組成物を投与することを含むことを特徴とする治療方法。

30

【請求項 19】

請求項 1 に記載のポリペプチドのアゴニストとして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 のポリペプチドを含むサンプルを化合物に曝露するステップと、

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

【請求項 20】

請求項 19 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成分。

40

【請求項 21】

機能的な ATR S の発現の低下に関連する疾患や病態の治療方法であって、そのような治療が必要な患者に請求項 20 の組成物を投与することを含むことを特徴とする治療方法。

【請求項 22】

請求項 1 に記載のポリペプチドのアンタゴニストとして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 のポリペプチドを含むサンプルを化合物に曝露するステップと、

(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ステップとを含むことを特徴とするスクリーニング方法。

【請求項 23】

50

請求項 22 に記載の方法によって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含むことを特徴とする成分。

【請求項 24】

機能的な A T R S 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、そのような治療が必要な患者に請求項 23 の組成物を投与することを含むことを特徴とする治療方法。

【請求項 25】

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スクリーニングする方法であって、

(a) 適切な条件下で請求項 1 に記載のポリペプチドを少なくとも 1 つの試験化合物に結合させる過程と、

(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 26】

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の活性が許容された条件下で、請求項 1 に記載のポリペプチドを少なくとも 1 つの試験化合物に結合させる過程と、

(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 27】

請求項 5 の配列を含む標的ポリヌクレオチドの発現を変化させ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、

(a) 前記標的ポリヌクレオチドの発現に好適な条件下で、前記標的ポリヌクレオチドを含むサンプルを化合物に曝露するステップと、

(b) 前記標的ポリヌクレオチドの変異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 28】

試験化合物の毒性を算定する方法であって、

(a) 核酸を含む生物学的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と、請求項 11 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を含むプローブをハイブリダイズさせるステップであって、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標的ポリヌクレオチドとの間で特異的なハイブリダイゼーション複合体が形成される条件下で行われ、前記標的ポリヌクレオチドが、請求項 11 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含むポリヌクレオチドである、前記ステップと、

(c) ハイブリダイゼーション複合体の収量を定量するステップと、

(d) 前記処理された生物学的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、処理されていない生物学的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、前記処理された生物学的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 29】

生物学的サンプル中の A T R S 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条件下で、前記生物学的サンプルを請求項 10 に記載の抗体と結合する過程

10

20

30

40

50

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、前記生物学的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 30】

前記抗体が、

(a) キメラ抗体

(b) 単鎖抗体

(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請求項 10 に記載の抗体。

10

【請求項 31】

請求項 10 に記載の抗体と、許容できる賦形剤とを含む化合物。

【請求項 32】

被検者の A T R S の発現に関連する病状又は疾患の診断方法であって、請求項 31 に記載の化合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 33】

前記抗体が標識されることを特徴とする請求項 31 に記載の化合物。

【請求項 34】

被検者の A T R S の発現に関連する病状又は疾患の診断方法であって、請求項 33 に記載の化合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。

20

【請求項 35】

請求項 10 に記載の抗体の特異性を有するポリクローナル抗体を調製する方法であって、(a) 抗体反応を誘発する条件下で、SEQ ID NO: 1 - 4 を有する群から選択したアミノ酸配列含むポリペプチドまたはその免疫抗原性断片を用いて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ポリペプチドでスクリーニングし、それによって、SEQ ID NO: 1 - 4 を有す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 36】

30

請求項 35 に記載の方法で産出した抗体。

【請求項 37】

請求項 36 に記載の抗体及び適切なキャリアを含む化合物。

【請求項 38】

請求項 10 に記載の抗体の特異性を有する抗体を用いてモノクローナル抗体を製造する方法であって、

(a) 抗体反応を誘発する条件下で、SEQ ID NO: 1 - 4 を有する群から選択したアミノ酸配列含むポリペプチドまたはその免疫抗原性断片を用いて動物を免疫化する過程と、

(b) 前記動物から抗体産出細胞を単離する過程と、

40

(c) 不死化の細胞を用いて前記抗体産出細胞を融合して、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 NO: 1 - 4 を有す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前記培養モノクローナル抗体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 39】

請求項 38 に記載の方法で産出したモノクローナル抗体。

【請求項 40】

請求項 39 に記載の抗体及び適切なキャリアを含む化合物。

50

【請求項 4 1】

F a b 発現ライブラリのスクリーニングにより前記抗体を産出することを特徴とする請求項 1 0 に記載の抗体。

【請求項 4 2】

組換え免疫グロブリンライブラリのスクリーニングにより前記抗体を産出することを特徴とする請求項 1 0 に記載の抗体。

【請求項 4 3】

S E Q I D N O : 1 - 4 を有す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検出する方法であって、

(a) 前記抗体と前記ポリペプチドの特異結合を許容する条件下で、サンプルを用いて請求項 1 0 に記載の抗体をインキュベートする過程と、 10

(b) 特異結合を検出する過程とを含み、該特異結合が、S E Q I D N O : 1 - 4 を有す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 4 4】

S E Q I D N O : 1 - 4 を有す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の特異結合を許容する条件下で、サンプルを用いて請求項 1 0 に記載の抗体をインキュベートする過程と、

(b) 前記サンプルから前記抗体を分離し、S E Q I D N O : 1 - 4 を有する群から選択した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 20

【請求項 4 5】

S E Q I D N O : 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 4 6】

S E Q I D N O : 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 4 7】

S E Q I D N O : 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 4 8】

S E Q I D N O : 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。 30

【請求項 4 9】

S E Q I D N O : 5 のポリヌクレオチド配列を有する請求項 1 1 に記載のポリヌクレオチド。

【請求項 5 0】

S E Q I D N O : 6 のポリヌクレオチド配列を有する請求項 1 1 に記載のポリヌクレオチド。

【請求項 5 1】

S E Q I D N O : 7 のポリヌクレオチド配列を有する請求項 1 1 に記載のポリヌクレオチド。

【請求項 5 2】 40

S E Q I D N O : 8 のポリヌクレオチド配列を有する請求項 1 1 に記載のポリヌクレオチド。

【発明の詳細な説明】

【0 0 0 1】

(技術分野)

本発明は、アミノアシル t R N A 合成酵素の核酸配列及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細胞増殖異常、自己免疫/炎症性の疾患の診断・治療・予防に関する。本発明はさらに、アミノアシル t R N A 合成酵素の核酸配列及びアミノ酸配列の発現における外来性化合物の効果についての評価に関する。

【0 0 0 2】 50

(発明の背景)

遺伝子コードの正確な翻訳は、個々のアミノ酸が適切な転移 RNA (tRNA) に結合することに依存する。アミノアシル-tRNA 合成酵素 (aaRS) は全生物に発見される必須タンパク質である。aaRS は、タンパク質の生合成の第一段階として、アミノ酸の活性化と同族の tRNA への正しい結合に関与している。原核生物は少なくとも 20 の異なる種類の aaRS を有している。異なる各アミノ酸には 1 つの aaRS がある。その一方、真核生物は個々のアミノ酸に対して、通常、細胞質型、ミトコンドリア型と呼ばれる 2 つの aaRS タイプを有する。20 個の aaRS 酵素は 2 つの構造クラスに分割できる。クラス I 酵素はアミノ酸を tRNA の 3' 端の 2' ヒドロキシル基に加える。クラス II 酵素はアミノ酸を tRNA の 3' 端の 3' ヒドロキシル基に加える。各クラスは触媒ドメインの特有のトポロジーによって特徴付けられる。クラス I 酵素は、ヌクレオチド結合部位の「Rossmann fold」にもとづいた触媒ドメインを有する。特に、コンセンサス テトラペプチド モチーフは高度に保存される (Prosite ドキュメント PDOC00161、アミノアシル移転 RNA 合成酵素クラス - I シグネチャ)。クラス I 酵素はアルギニン、システイン、グルタミン酸、グルタミン、イソロイシン、ロイシン、メチオニン、チロシン、トリプトファンおよびバリンに対して特異的である。クラス II 酵素は中央触媒ドメインを含み、7 つのストランドをもつ反パラレル β -シート ドメイン、及び N 末端 - と C 末端の調節ドメインから構成される。クラス II 酵素は酵素がヘテロ 2 量体構造かまたはホモ 2 量体構造かによって 2 つのグループに分かれる。後者のグループは、N 末端および C 末端の調節ドメイン (Hartle in, M. 及び Cusack, S. (1995) J. Mol. Evol. 40: 519-530) の構造によってさらに分かれる。クラス II 酵素はアラミン、アスパラギン、アスパラギン酸、グリシン、ヒスチジン、リシン、フェニルアラニン、プロリン、セリン、トレオニンに特定されている。

【0003】

特定の aaRS には編集 (エディティング) 機能もある。たとえば、IleRS は Val-tRNA^{Ile} を形成して、バリンを誤って活性化させる場合があるが、この生成物は誤って結合した生成物を破壊する加水分解活性によって取り除かれる。この編集活性は、クラス I 酵素の Rossmann fold ドメイン内の長い挿入配列である接続ポリペプチド 1 領域 (CP1) で見つかる 2 番目の触媒サイト内に位置する。(Schimmel, P. 他 (1998) FASEB J. 12: 1599-1609)。AaRS はまた tRNA プロセッシングでも役割を果たす。成熟した tRNA は細胞質に移出される前に、核の中のそれぞれのアミノ酸と結合することが示されている。またこのアミノ酸との結合は品質管理機構として機能し、tRNA が正常に機能するようにする (Martinis, S.A. 他 (1999) EMBO J. 18: 4591-4596)。

【0004】

タンパク質合成における機能に加えて、特定のアミノアシル tRNA 合成酵素は細胞忠実性、RNA スプライシング、RNA 輸送、アポトーシスおよび転写と翻訳の調節で役割を果たす。たとえば、ヒトのチロシル tRNA 合成酵素はたんぱく質分解により、特有のシトキン活性を持つ 2 つの部分に裂ける。C 末端ドメインは単球と白血球の走化性を示し、また、ミエロペルオキシダーゼ、腫瘍壊死因子 - α および組織因子の生成を促す。N-末端ドメインはインターロイキン - 8 タイプ A 受容体に結合し、インターロイキン 8 様シトキンとして機能する。ヒトのチロシル-tRNA 合成酵素はアポトーシス腫瘍細胞から分泌され、アポトーシスを加速することがある (Wakasugi, K および Schimmel, P (1999) Science 284: 147-151)。アカパンカビ (*Neurospora crassa*) のミトコンドリア TyrRS および出芽酵母 (*S. cerevisiae*) の LeuRS は特定のグループ I イントロンのスプライシング活動に必須であり、ヒトのミトコンドリア LeuRS を酵母のヌル種 (null strain) において酵母の LeuRS として代用にできる

。特定の バクテリアの a a R S はそれ自身の転写または翻訳の調節に関与している（前出 Martini s）。一部の a a R S は、細胞増殖、分化、およびアポトーシスで役割をもったシグナル分子のクラスである、ジアデノシン オリゴホスフェートを合成できる（Kissel ev, L. L 他（1998）F E B S L e t t . 4 2 7 : 1 5 7 - 1 6 3 ; V a r t a n i a n , A . e t a l . (1 9 9 9) F E B S L e t t . 4 5 6 : 1 7 5 - 1 8 0 ）。

【0005】

アミノアシル - t R N A への自己抗体は、リウマチ関節炎、皮膚筋炎、及び多発性筋炎などの自動免疫症のある患者によって生成され、併発する間質性肺疾患（I L D）に強度の関連を持つ（Freist, W. 他（1999）Bi ol . Chem . 3 8 0 : 6 2 3 - 6 4 6 ; Freist, W. 他（1996）Bi ol . Chem . H o p p e S e y l e r 3 7 7 : 3 4 3 - 3 5 6）。これらの自己抗体はウィルス感染への反応として生成され、コクサッキー ウィルスは動物に実験的ウィルス性筋肉炎を誘発する。

10

【0006】

a a R S 構造の人体と病原体との比較は新規の抗生物質の設計に有効である（Schimmel, 前出）。遺伝子工学で作られた a a R S は、自然にないアミノ酸を部位特異的にたんぱく質へ i n v i v o で取り込ませるために利用されている（Liu, D. R. 他（1997）Proc . Nat l . Acad . Sci . USA 9 4 : 1 0 0 9 2 - 1 0 0 9 7）。

20

新規のアミノアシル t R N A 合成酵素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、細胞増殖異常、自己免疫 / 炎症性の疾患の診断・治療・予防において有用であり、また、アミノアシル t R N A 合成酵素の核酸配列及びアミノ酸配列の発現における外来性化合物の効果についての評価にも有用である。

【0007】

（発明の概要）

本発明の機能は、精製されたポリペプチドである集合的に「A T R S」と呼ばれ、個別には「A T R S - 1」、「A T R S - 2」、「A T R S - 3」および「A T R S - 4」であるアミノアシル t R N A 合成酵素である。或る実施態様において本発明は、（a）S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列からなるポリペプチド、（b）S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列と少なくとも90%が同一であるアミノ酸配列を有する天然のポリペプチド、（c）S E Q I D N O : 1 - 4 を有す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、または（d）S E Q I D N O : 1 - 4 を有す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抗原性断片を含む群から選択した実質上単離されたポリペプチドを提供する。一実施態様では、S E Q I D N O : 1 - 4 のアミノ酸配列を含む実質上単離されたポリペプチドを提供する。

30

【0008】

また、本発明は（a）S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列、（b）S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列と少なくとも90%の相同性を有する天然のアミノ酸配列、（c）S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列の生物学的活性断片、または（d）S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列の免疫抗原性断片を含むポリペプチドをコードするような実質上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、ポリヌクレオチドはS E Q I D N O : 1 - 4 を有する群から選択した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S E Q I D N O : 5 - 8 を有する群から選択される。

40

【0009】

本発明は更に、（a）S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列か

50

らなるポリペプチド、(b)SEQ ID NO: 1-4を有する群から選択したアミノ酸配列と少なくとも90%の相同性を有するアミノ酸配列からなる天然のポリペプチド、(c)SEQ ID NO: 1-4を有す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、または(d)SEQ ID NO: 1-4を有す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抗原性断片から構成される群から選択された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結合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を提供する。一実施態様では、本発明は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。別の実施態様では、本発明は組換えポリヌクレオチドを含む遺伝形質転換体を提供する。

【0010】

10

また、本発明は(a)SEQ ID NO: 1-4を有する群から選択したアミノ酸配列、(b)SEQ ID NO: 1-4を有する群から選択したアミノ酸配列と少なくとも90%の相同性を有する天然のアミノ酸配列、(c)SEQ ID NO: 1-4を有する群から選択したアミノ酸配列の生物学的活性断片、または(d)SEQ ID NO: 1-4を有する群から選択したアミノ酸配列の免疫抗原性断片を含む実質上単離されたポリペプチドを製造する方法を提供する。製造方法は、(a)組換えポリヌクレオチドを用いて形質転換した細胞をポリペプチドの発現に適した条件下で培養する過程と、(b)そのように発現したポリペプチドを受容する過程とを有し、組換えポリヌクレオチドはポリペプチドをコードするポリヌクレオチドに機能的にリンクしたプロモーター配列を有する。

【0011】

20

本発明は更に、(a)SEQ ID NO: 1-4を有する群から選択したアミノ酸配列を有するポリペプチド、(b)SEQ ID NO: 1-4を有する群から選択したアミノ酸配列と少なくとも90%同一であるアミノ酸配列を有する天然のポリペプチド、(c)SEQ ID NO: 1-4を有す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、または(d)SEQ ID NO: 1-4を有す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抗原性断片から構成される群から選択されたポリペプチドに特異結合するような実質上単離された抗体を提供する。

【0012】

本発明は更に、(a)SEQ ID NO: 5-8を有する群から選択したポリヌクレオチド配列、(b)SEQ ID NO: 5-8を有す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、(c)(a)に相補的なポリヌクレオチド配列、(d)(b)に相補的なポリヌクレオチド配列、または(e)(a)~(d)のRNA等価物を含む実質上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、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を有する。

30

【0013】

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: 5-8を有する群から選択したポリヌクレオチド配列、(b)SEQ ID NO: 5-8を有す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、(c)(a)に相補的なポリヌクレオチド配列、(d)(b)に相補的なポリヌクレオチド配列、または(e)(a)~(d)のRNA等価物を含む実質上単離されたポリヌクレオチドを提供する。検出方法は、(a)サンプル中の標的ポリヌクレオチドに相補的な配列からなる少なくとも20の連続したヌクレオチドを含む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)ハイブリダイゼーション複合体の存在・不存在を検出し、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からなり、プローブと標的ポリヌクレオチド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、プローブは標的ポリヌクレオチドに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を含む。

40

【0014】

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

50

、標的ポリヌクレオチドは (a) S E Q I D N O : 5 - 8 を有する群から選択したポリヌクレオチド配列、 (b) S E Q I D N O : 5 - 8 を有す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、 (c) (a) に相補的なポリヌクレオチド配列、 (d) (b) に相補的なポリヌクレオチド配列、または (e) (a) ~ (d) の R N A 等価物を含む実質上単離されたポリヌクレオチドを提供する。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) 標的ポリヌクレオチドまたはその断片の存在・不存在を検出し、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を含む。

【 0 0 1 5 】

10

本発明は更に、有効量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成分を提供する。有効量のポリペプチドは、 (a) S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列、 (b) S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列と少なくとも 90 % の相同性を有する天然のアミノ酸配列、 (c) S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列の生物学的活性断片、または (d) S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列の免疫抗原性断片を含む。一実施例では、S E Q I D N O : 1 - 4 からなる一群から選択されたアミノ酸配列を含む組成物を提供する。更に、本発明は、患者にこの組成物を投与することを含む、機能的 A T R S の発現の低下に関連した疾患やその症状の治療方法を提供する。

【 0 0 1 6 】

20

本発明はまた、 (a) S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列、 (b) S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列と少なくとも 90 % の相同性を有する天然のアミノ酸配列、 (c) S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列の生物学的活性断片、または (d) S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列の免疫抗原性断片を含む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を提供する。スクリーニング方法は、 (a) ポリペプチドを有するサンプルを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とを含む。別法では、本発明は、この方法によって同定されたアゴニスト化合物と好適な医薬用賦形剤とを含む組成物を提供する。更なる別法では、本発明は、この組成物の患者への投与を含む、機能的 A T R S の発現の低下に

30

【 0 0 1 7 】

本発明は更に、 (a) S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列、 (b) S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列と少なくとも 90 % の相同性を有する天然のアミノ酸配列、 (c) S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列の生物学的活性断片、または (d) S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列の免疫抗原性断片を含む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を提供する。スクリーニング方法は、 (a) ポリペプチドを含むサンプルを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とを含む。一実施態様で本発明は、この方法によって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含む成分を提供する。更なる別法では、本発明は、このような治療を必要とする患者へのこの組成物の投与を含む、機能的 A T R S の過剰な発現に関連した疾患やその症状の治療方法を提供する。

40

【 0 0 1 8 】

本発明は更に、 (a) S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列、 (b) S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列と少なくとも 90 % の相同性を有する天然のアミノ酸配列、 (c) S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列の生物学的活性断片、または (d) S E Q I D N O : 1 - 4 を有する群から選択したアミノ酸配列の免疫抗原性断片を含むポリペプチドに特異結合する

50

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)ポリペプチドを適切な条件下で少なくとも1つの試験化合物に結合させる過程と、(b)試験化合物とのポリペプチドの結合を検出し、それによって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。

【0019】

この方法は、(a)SEQ ID NO: 1 - 4を有する群から選択したアミノ酸配列、(b)SEQ ID NO: 1 - 4を有する群から選択したアミノ酸配列と少なくとも90%の相同性を有する天然のアミノ酸配列、(c)SEQ ID NO: 1 - 4を有する群から選択したアミノ酸配列の生物学的活性断片、または(d)SEQ ID NO: 1 - 4を有する群から選択したアミノ酸配列の免疫抗原性断片を含むポリペプチドの活性を調節す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)ポリペプチドの活性が許容された条件下で、ポリペプチドを少なくとも1つの試験化合物と混合させる過程と、(b)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ポリペプチドの活性を試験化合物の不存在下で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ポリペプチドの活性の変化は、ポリペプチドの活性を調節する化合物であることを意味する。

10

【0020】

本発明は更に、標的ポリヌクレオチドの変異発現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を提供する。標的ポリヌクレオチドは、SEQ ID NO: 5 - 8を有する群から選択した配列を含む。スクリーニング方法は、(a)標的ポリヌクレオチドを含むサンプルを化合物に曝す過程と、(b)標的ポリヌクレオチドの変異発現を検出する過程とを含む。

20

【0021】

本発明は更に、(a)核酸を含む生物学的サンプルを試験化合物で処理する過程と、(b)(i)SEQ ID NO: 5 - 8を有す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、(ii)SEQ ID NO: 5 - 8を有す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%の相同性を有するポリヌクレオチド配列を含む天然のポリヌクレオチド、(iii)(i)に相補的な配列を有するポリヌクレオチド、(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等価物を含む群から選択した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から構成されるプローブを用いて、処理した生物学的サンプルの核酸をハイブリダイズする過程とを含む試験化合物の毒性の算定方法を提供する。ハイブリダイゼーションは、上記プローブと生物学的サンプル中の標的ポリヌクレオチドの間に特定の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発生し、上記標的ポリヌクレオチドは、(i)SEQ ID NO: 5 - 8を有す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、(ii)SEQ ID NO: 5 - 8を有す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%の相同性を有するポリヌクレオチド配列を含む天然のポリヌクレオチド、(iii)(i)のポリヌクレオチドに相補的な配列を有するポリヌクレオチド、(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等価物を含む群から選択する。或いは、標的ポリヌクレオチドは、上記(i) ~ (v)を含む群から選択したポリヌクレオチド配列の断片と、(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)処理した生物学的サンプルのハイブリダイゼーション複合体の量を、非処理の生物学的サンプルのハイブリダイゼーション複合体の量と比較する過程を含み、処理した生物学的サンプルのハイブリダイゼーション複合体の量の差は、試験化合物の毒性を示す。

30

40

【0022】

(発明を実施するための形態)

本発明のタンパク質、ヌクレオチド配列及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、材料及び方法に本発明が限定されるものではなく、改変し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例を説明する目的で用いたものに

50

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あることも併せて理解されたい。

【0023】

請求の範囲及び明細書中で用いている単数形の「或る」及び「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しなければならない。従って、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単数または複数の抗体、及び、当業者に公知の抗体の等価物等についても言及しているのである。

【0024】

本明細書中で用いる全ての専門用語及び科学用語は、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、材料及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があるであろう細胞、プロトコル、試薬及びベクターについて説明及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。

【0025】

(定義)

用語「ATRS」は、天然、合成、半合成或いは組換え体など全ての種(特にウシ、ヒツジ、ブタ、マウス、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ATRSのアミノ酸配列を指す。

【0026】

用語「アゴニスト」は、ATRSの生物学的活性を強めたり、模倣する分子を指す。このアゴニストは、ATRSに直接相互作用するか、或いはATRSが関与する生物学的経路の成分と作用して、ATRSの活性を調節するタンパク質、核酸、糖質、小分子、任意の他の化合物や組成物を含み得る。

【0027】

用語「対立遺伝子変異配列」は、ATRSをコードする遺伝子の別の形を指す。対立遺伝子変異体は、核酸配列における少なくとも1の突然変異から作製し得る。また、変異RNAまたはポリペプチドからも作製し得る。ポリペプチドの構造または機能は、変異することもしないこともある。遺伝子は、天然の対立遺伝子変異体を全く有しないか、1若しくは数個の天然の対立遺伝子変異体を有し得る。一般に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、ヌクレオチドの自然欠失、付加または置換に帰するものである。これら各変化は、単独或いは他の変化と共に、所定の配列内で1若しくは数回生じ得る。

【0028】

ATRSをコードする「変異」核酸配列は、様々なヌクレオチドの欠失、挿入、或いは置換が起こっても、ATRSと同じポリペプチド或いはATRSの機能特性の少なくとも1つを備えるポリペプチドを指す。この定義には、ATRSをコードするポリヌクレオチド配列の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配列との不適當或いは予期しないハイブリダイゼーション、並びにATRSをコードするポリヌクレオチドの特定の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形性を含む。コードされたタンパク質も変異され得り、サイレント変化を生じATRSと機能的に等価となるアミノ酸残基の欠失、挿入、或いは置換を含み得る。意図的なアミノ酸置換は、生物学的或いは免疫学的にATRSの活性が保持される範囲で、残基の極性、電荷、溶解度、疎水性、親水性、及び/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及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及びアルギニンがある。親水性値が近似している非荷電極性側鎖を有するアミノ酸には、アスパラギンとグルタミン、セリンとスレオニンがある。親水性値が

10

20

30

40

50

近似している非荷電側鎖を有するアミノ酸には、ロイシンとイソロイシンとバリン、グリシンとアラニン、フェニルアラニンとチロシンがある。

【 0 0 2 9 】

「アミノ酸」または「アミノ酸配列」の語は、オリゴペプチド、ペプチド、ポリペプチド若しくはタンパク質の配列またはその断片を指し、天然または合成分子を指す。ここで、「アミノ酸配列」は天然のタンパク質分子のアミノ酸配列を指すものであり、「アミノ酸配列」及び類似の語は、アミノ酸配列を、列挙したタンパク質分子に会合する完全な本来のアミノ酸配列に限定しようとするものではない。

【 0 0 3 0 】

「増幅」は、核酸配列の追加複製に関連する。増幅は通常、当業者によく知られたポリメラーゼ連鎖反応 (P C R) 技術を用いて行う。 10

【 0 0 3 1 】

用語「アンタゴニスト」は、A T R S の生物学的活性を阻害或いは減弱する分子である。アンタゴニストは、A T R S に直接相互作用するか、或いはA T R S が関与する生物学的経路の成分と作用して、A T R S の活性を調節する抗体、核酸、糖質、小分子、任意の他の化合物や組成物などのタンパク質を含み得る。

【 0 0 3 2 】

「抗体」の語は、無損傷免疫グロブリンやその断片、例えばF a、F (a b ')₂ 及びF v断片を指すが、これらはエピトープの決定基と結合することができる。A T R S ポリペプチドと結合する抗体は、抗体を免疫する小ペプチドを含む無傷の分子またはその断片を用いて作製可能である。動物 (マウス、ラット、ウサギ等) 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は、翻訳または化学合成されたR N A に由来し得るので、好みに応じて担体タンパク質に接合することも可能である。通常用いられる担体であってペプチドと化学結合するものは、ウシ血清アルブミン、サイログロブリン及びキーホールリンペットヘモシアニン (K L H) 等がある。結合ペプチドは、動物を免疫化するために用いる。 20

【 0 0 3 3 】

「抗原決定基」の語は、特定の抗体と接触している分子の領域 (即ちエピトープ) 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導し得る。抗原決定基は、抗体に結合するための無損傷抗原 (即ち免疫応答を誘導するために用いられる免疫原) と競合し得る。 30

【 0 0 3 4 】

「アンチセンス」の語は、特定の核酸配列の「センス」鎖と塩基対を形成することが可能な任意の成分を指す。アンチセンス成分には、D N A や、R N A や、ペプチド核酸 (P N A) や、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等の修飾されたバックボーン連鎖を有するオリゴヌクレオチドや、2' - メトキシエチル糖または2' - メトキシエトキシ糖等の修飾された糖類を有するオリゴヌクレオチドや、或いは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等の修飾された塩基を有するオリゴヌクレオチドがある。アンチセンス分子は、化学合成または転写を含む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、ひとたび細胞に導入されたら、細胞が形成した天然の核酸配列と塩基対を形成し、転写または翻訳を妨害する二重鎖を形成する。「負」若しくは「マイナス (-) 」の語がアンチセンス鎖を、「正」若しくは「プラス (+) 」がセンス鎖を指すことがある。 40

【 0 0 3 5 】

「生物学的に活性」の語は、天然分子の構造的機能、調節機能または生化学的機能を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、天然或いは組換え体のA T R S、合成のA T R S 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。

【 0 0 3 6 】

「相補(的)」または「相補性」の語は、ポリヌクレオチドとポリヌクレオチドの、塩基対形成による自然結合を指す。例えば、配列「5' A - G - T 3'」は、相補配列「3' T - C - A 5'」に結合する。

【0037】

「所定のポリヌクレオチド配列からなる成分」及び「所定のアミノ酸配列からなる成分」は、概して所定のポリヌクレオチド配列またはアミノ酸配列からなる任意の成分を指す。この成分には、乾燥製剤または水溶液が含まれ得る。A T R S もしくはA T R S の断片をコードするポリヌクレオチド配列を含む組成物は、ハイブリダイゼーションプローブとして使用され得る。このプローブは、凍結乾燥状態で保存可能であり、糖質などの安定化剤と結合させることが可能である。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)及びその他の構成エレメント(例えばデンハート液、脱脂粉乳、サケの精子のDNA等)を含む水溶液中にプローブを分散させることができる。

10

【0038】

「コンセンサス配列」は、不要な塩基を分離するためにDNA配列の解析を繰り返し行い、XL-PCRキット(PE Biosystems, Foster City CA)を用いて5' 及び/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断片構築システム(GCG, Madison, WI)またはPhrap(University of Washington, Seattle WA)等の断片構築用のコンピュータプログラムを用いて1つ或いはそれ以上の重複するcDNAやEST、またはゲノムDNA断片から構築された核酸配列を指す。伸長及びアセンブルの両方を行ってコンセンサス配列を決定する配列もある。

20

【0039】

「保存的なアミノ酸置換」は、置換がなされた時に元のタンパク質の特性を殆ど損なわないような置換、即ちタンパク質の構造と特に機能が保存され、そのような置換による大きな変化がない置換を指す。下表は、タンパク質中で元のアミノ酸と置換され得るアミノ酸と、保存アミノ酸置換と認められるアミノ酸を示している。

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	10
Gln	Asn, Glu, His	
Glu	Asp, Gln, His	
Gly	Ala	
His	Asn, Arg, Gln, Glu	
Ile	Leu, Val	
Leu	Ile, Val	20
Lys	Arg, Gln, Glu	
Met	Leu, Ile	
Phe	His, Met, Leu, Trp, Tyr	
Ser	Cys, Thr	
Thr	Ser, Val	
Trp	Phe, Tyr	30

【 0 0 4 0 】

保存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 シートや ヘリックス構造、(b) 置換部位における分子の電荷または疎水性、及び / または (c) 側鎖の大部分を保持する。

【 0 0 4 1 】

「欠失」は、結果的に 1 若しくは数個のアミノ酸またはヌクレオチドが失われてなくなるようなアミノ酸またはヌクレオチド配列における変化を指す。

【 0 0 4 2 】

「誘導体」の語は、ポリペプチド配列またはポリヌクレオチド配列の化学修飾を指す。例えば、アルキル基、アシル基、ヒドロキシシル基またはアミノ基による水素の置換は、ポリヌクレオチド配列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、グリコシル化、ポリエチレングリコール化 (p e g y l a t i o n)、或いは任意の同様なプロセスであって誘導起源のポリペプチドから少なくとも 1 つの生物学的若しくは免疫学的機能を保持しているプロセスによって、修飾されたポリペプチドである。

【 0 0 4 3 】

「検出可能な標識」は、測定可能な信号を生成することができ、ポリヌクレオチドまたは

10

20

30

40

50

ポリペプチドに共有結合または非共有結合するようなレポーター分子または酵素を指す。

【0044】

「示差発現」は少なくとも2つの異なったサンプルを比較することによって決められる、増加、または非調節、あるいは減少、下方調節、または欠損遺伝子またはタンパク発現を指す。このような比較は例えば、治療後サンプルと未治療のサンプルまたは病態のサンプルと正常サンプルの間で行われ得る。

【0045】

用語「断片」は、A T R S または A T R S 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 (p a r e n t s e q u e n c e) と同一であるがその配列より長さが短いものを指す。断片は、画定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、或いはその他の目的のために用いられる断片は、少なくとも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或いはアミノ酸残基長さであり得る。断片は、分子の特定領域から優先的に選択し得る。例えば、ポリペプチド断片は、所定の配列に示すような最初の250または500アミノ酸（またはポリペプチドの最初の25%または50%）から選択された或る長さの連続したアミノ酸を有し得る。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、本発明の実施例では、配列表、表及び図面を含む明細書に裏付けされた任意の長さであってよい。

10

20

【0046】

S E Q I D N O : 5 - 8 の断片には、固有のポリヌクレオチド配列領域が含まれる。この領域は、S E Q I D N O : 5 - 8 を特異的に同定するものであり、例えば同一ゲノム中の S E Q I D N O : 5 - 8 以外の配列とは異なるものである。S E Q I D N O : 5 - 8 の断片は、例えば、ハイブリダイゼーション及び増幅技術において、或いは関連するポリヌクレオチド配列から S E Q I D N O : 5 - 8 を区別する類似の方法において有用である。ある断片と一致する S E Q I D N O : 5 - 8 の正確な断片の長さや領域は、その断片の目的に基づいて当分野で一般的な技術によって日常的に測定できる。

【0047】

S E Q I D N O : 1 - 4 の断片は、S E Q I D N O : 5 - 8 の断片によってコードされる。S E Q I D N O : 1 - 4 の断片には、S E Q I D N O : 1 - 4 を特異的に同定する固有のアミノ酸配列領域が含まれている。例えば、S E Q I D N O : 1 - 4 の断片は、S E Q I D N O : 1 - 4 を特異認識する抗体を産出するための免疫抗原性ペプチドとして有用である。S E Q I D N O : 1 - 4 の断片及び断片に対応する S E Q I D N O : 1 - 4 の領域の正確な長さは、断片に対する意図した目的に基づき当業者が慣例的に決定することが可能である。

30

【0048】

「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）、オープンリーディングフレーム及び翻訳終止コドンを有する配列である。「完全長」ポリヌクレオチド配列は、「完全長」ポリペプチド配列をコードする。

40

【0049】

「相同性」の語は、配列類似性即ち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間で互換可能な配列同一性である。

【0050】

ポリヌクレオチド配列に適用される「一致率」または「一致%」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少なくとも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。このようなアルゴリズムは、2配列間のアラインメントを最適化するために比較する配列において、標準化された再現性のある方法でギャップを挿入するので、2つの配列をより有意に比較できる。

【0051】

50

ポリヌクレオ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を用いて決定できる。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. 及び P. M. Sharp (1989) CABIOS 5:151-153、Higgins, D. G. 他 (1992) CABIOS 8:189-191に記載されている。ポリペプチド配列を2つ1組で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、「diagonal saved」= 4と設定する。残基「重み付け」表をデフォルトとして選択する。CLUSTAL Vは、アラインメントされたポリヌクレオチド配列対間の「類似率」として一致率を報告する。

10

【0052】

或いは、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、無料で入手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. ら (1990) J. Mol. Biol. 215:403-410）。このアルゴリズムは、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及びインターネット（<http://www.ncbi.nlm.nih.gov/BLAST/>）からも入手可能である。BLASTソフトウェア一式には様々な配列分析プログラムが含まれており、既知のポリヌクレオチド配列を種々のデータベースから得た別のポリヌクレオチド配列とアラインメントする「blastn」もその1つである。その他にも、2つのヌクレオチド配列を対で直接比較するために用いる「BLAST 2 Sequences」と称されるツールも利用可能である。「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することが可能である。「BLAST 2 Sequences」ツールは、blastn と blastp（後述）の両方に用いることができる。BLASTプログラムは、一般的には、ギャップ及びデフォルト設定に設定された他のパラメータと共に用いる。例えば、2つのヌクレオチド配列を比較するために、デフォルトパラメータとして設定された「BLAST 2 Sequences」ツールVersion 2.0.9（1999年5月7日）を用いてblastnを実行してもよい。デフォルトパラメータの設定例を以下に示す。

20

30

【0053】

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

40

一致率は、完全に画定された（例えば特定の配列番号で画定された）配列長さと比較して測定し得る。或いは、より短い長さ、例えばより大きな画定された配列から得られた断片（例えば少なくとも20、30、40、50、70、100または200の連続したヌクレオチドの断片）の長さと比較して一致率を測定してもよい。ここに挙げた長さは単なる例示的なものに過ぎず、表、図及び配列リストを含めた本明細書に記載された配列に裏付けられた任意の配列長さの断片を用いて、一致率を測定し得る長さを説明し得ることを理解されたい。

【0054】

高度の相同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

50

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。

【0055】

ポリペプチド配列に適用される「一致率」または「一致%」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少なくとも2以上のポリペプチド配列間で一致する残基の割合を意味する。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、置換部位の酸性度及び疎水性を保存するので、ポリペプチドの構造を（従って機能も）保存する。

10

【0056】

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を用いて決定できる（既に説明したのでそれを参照されたい）。CLUSTAL V を用いて、ポリペプチド配列を2つ1組で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 と設定する。デフォルトの残基重み付け表としてPAM250マトリクスを選択する。ポリヌクレオチドアラインメントと同様に、CLUSTAL V は、アラインメントされたポリペプチド配列対間の「類似率」として一致率を報告する。

20

【0057】

或いは、NCBI BLAST ソフトウェア一式を用いてもよい。例えば、2つのポリペプチド配列を対で比較をする場合、ある者は、デフォルトパラメータで設定された「BLAST 2 Sequences」ツール Version 2.0.12 (Apr - 21 - 2000) でblastpを使用するであろう。デフォルトパラメータの設定例を以下に示す。

【0058】

Matrix: BLOSUM62
Open Gap: 11 及び Extension Gap: 1 penalties
Gap x drop-off: 50
Expect: 10
Word Size: 3
Filter: on

30

一致率は、完全に画定された（例えば特定の配列番号で画定された）ポリペプチド配列の長さと比較して測定し得る。一致率は、配列或いは、より短い長さ、例えばより大きな画定されたポリペプチド配列から得られた断片（例えば少なくとも15、20、30、40、50、70または150の連続した残基の断片）の長さと比較して一致率を測定してもよい。ここに挙げた長さは単なる例示的なものに過ぎず、表、図及び配列リストを含めた本明細書に記載された配列に裏付けられた任意の配列長さの断片を用いて、或る長さであってその長さに対して一致率を測定し得る長さを説明し得ることを理解されたい。

40

【0059】

「ヒト人工染色体（HAC）」は、約6kb（キロベース）～10MbのサイズのDNA配列を含み得る、安定した染色体複製の分離及び維持に必要な全てのエレメントを含む直鎖状の小染色体である。

【0060】

「ヒト化抗体」の語は、非抗体結合領域におけるアミノ酸配列はヒト抗体により近づくように変異させた抗体分子であって、本来の結合能力はそのまま保持しているような抗体分子を指す。

【0061】

「ハイブリダイゼーション」は、所定のハイブリダイゼーション条件下で塩基対を形成す

50

ることによって、一本鎖ポリヌクレオチドが相補的鎖とアニーリングするプロセスを指す。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同性を共有することを示すものである。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、「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、更にストリンジェントな条件では、非特異結合（即ち完全には一致しない核酸鎖間の対の結合）が減少する。核酸配列のアニーリングに対する許容条件は、本技術分野における当業者が慣例的に決定する。許容条件はハイブリダイゼーション実験の間は一定でよいが、洗浄条件は所望のストリンジェンシーを得るように、従ってハイブリダイゼーション特異性も得るように実験中に変更することができる。アニーリングが許容される条件は、例えば、温度が68℃で、約6×SSC、約1%（w/v）のSDS、並びに約100μg/mlのせん断して変性したサケ精子DNAが含まれる。

10

【0062】

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、所定のイオン強度及びpHにおける特異配列の融点（ T_m ）より約5～20℃低くなるように選択する。この T_m は、所定のイオン強度及びpHの下で、完全に一致する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 を計算する式及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ら（1989）Molecular Cloning: 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。

20

【0063】

本発明のポリヌクレオチド間の高いストリンジェンシーのハイブリダイゼーションでは、約0.2×SSC及び約1%のSDSの存在の下、約68℃で1時間の洗浄過程を含む。別法では、65℃、60℃、55℃、42℃の温度で行う。SSC濃度は、約0.1%のSDS存在下で、約0.1～2×SSCの範囲で変化し得る。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。このような遮断剤には、例えば、約100～200μg/mlの変性サケ精子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に有機溶剤、例えば約35～50%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。このような類似性は、ヌクレオチド及びヌクレオチドにコードされるポリペプチドに対する類似の役割を強く示唆している。

30

【0064】

「ハイブリダイゼーション複合体」の語は、相補的塩基対間の水素結合の形成力によって2つの核酸配列間に形成された複合体を指す。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（C₀tまたはR₀t解析等）。或いは、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、或いは他の適切な基質であって細胞若しくはその核酸が固定される基質）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。

40

【0065】

「挿入」及び「付加」の語は、1若しくは数個のアミノ酸残基またはヌクレオチド配列を各々付加するようなアミノ酸またはヌクレオチド配列における変化を指す。

【0066】

「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に関連する症状を指し得る。これらの症状は、細胞及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ることができる。

【0067】

50

用語「免疫原性断片」は、例えば哺乳動物などの生きている動物に導入すると、免疫反応を引き起こす A T R S 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 A T R S のポリペプチド断片またはオリゴペプチド断片を含む。

【0068】

「マイクロアレイ」の語は、基質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。

【0069】

「エレメント」または「アレイエレメント」の語は、マイクロアレイの環境において、基質の表面上に配置されたハイブリダイズ可能なポリヌクレオチドを指す。

10

【0070】

用語「調節」は、A T R S の活性の変化を指す。例えば、調節によって、A T R S のタンパク質活性、或いは結合特性、またはその他の生物学的特性、機能的特性或いは免疫学的特性の変化が起こる。

【0071】

「核酸」及び「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」及び「核酸配列」の語は、ゲノム起源または合成起源の D N A または R N A であって一本鎖または二本鎖であるか或いはセンス鎖またはアンチセンス鎖を表し得るような D N A または R N A や、ペプチド核酸 (P N A) や、任意の D N A 様または R N A 様物質を指すこともある。

20

【0072】

「機能的に結合した」は、第 1 核酸配列が第 2 核酸配列と機能的な関係があるように配置された状態を指す。例えば、プロモーターが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結合している。同一のリーディングフレーム内で 2 つのタンパク質コード領域を結合する必要がある場合、一般に、機能的に結合した D N A 配列は非常に近接するか、或いは連続し得る。

【0073】

「ペプチド核酸」(P N A) は、アンチセンス分子または抗遺伝子物質であって、リジン末端とするアミノ酸残基のペプチド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からなるものを指す。末端のリジンは、成分に溶解性を与える。P N A は、相補的一本鎖 D N A または R N A に優先的に結合して転写の拡張を停止するものであり、ポリエチレングリコール化して細胞における P N A の寿命を延長し得る。

30

【0074】

A T R S 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、蛋白分解性切断及びその他の当分野で既知の修飾を含まれ得る。これらのプロセスは、合成或いは生化学的に生じ得る。生化学的修飾は、A T R S の酵素環境に依存し、細胞の種類によって異なり得る。

【0075】

「プローブ」とは、同一配列或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、A T R S 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結合したものである。典型的な標識には、放射性アイソトープ、リガンド、化学発光試薬及び酵素がある。「プライマー」は、短い核酸、通常は D N A オリゴヌクレオチドであり、相補的塩基対を形成することで標的ポリヌクレオチドにアニーリングされ得る。プライマーは次に、D N A ポリメラーゼ酵素によって標的 D N A 鎖に延在し得る。プライマー対は、例えばポリメラーゼ連鎖反応 (P C R) による核酸配列の増幅 (及び同定) に用い得る。

40

【0076】

本発明に用いるようなプローブ及びプライマーは通常、既知の配列の少なくとも 15 の連

50

続したヌクレオチドを含んでいる。特異性を高めるために長めのプローブ及びプライマー、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150の連続したヌクレオチドからなるようなプローブ及びプライマーを用いてもよい。これよりもかなり長いプローブ及びプライマーもある。表、図面及び配列リストを含む本明細書に裏付けされた任意の長さのヌクレオチドを用いることができるものと理解されたい。

【0077】

プローブ及びプライマーの調製及び使用方法については、Sambrook, J. ら (1989) Molecular Cloning: A Laboratory Manual, 第2版, 1-3巻, Cold Spring Harbor Press, Plainview NY、Ausubel, F.M. ら、(1987) Current Protocols in Molecular Biology, Greene Publishing Assoc. & Wiley-Intersciences, New York NY、Innis ら (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA等を参照されたい。PCRプライマー対は、その目的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)を用いるなどして既知の配列から得ることができる。

【0078】

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及び最大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ものを分析するのににも有用である。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、従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(マサチューセッツ州ケンブリッジのWhitehead Institute/MITゲノム研究センターから一般向けに入手可能)ではユーザーが「ミスプライミング・ライブラリ」をインプットすることができ、ここでプライマー結合部位として避けたい配列はユーザーが指定する。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、各自のソースから得てユーザー固有のニーズを満たすように変更してもよい。)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何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。従って、このプログラムは、固有であって保存されたオリゴヌクレオチド及びポリヌクレオチドの断片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及びポリヌクレオチドの断片は、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、或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。

【0079】

「組換え核酸」は天然配列ではない配列であるか或いは人為的に組み合わせなければ分離しているような配列の2以上のセグメントを人為的に組み合わせで産出した配列を有する配列である。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

10

20

30

40

50

核酸の単離セグメントの人為的操作によって、例えばの S a m b r o o k らの文献（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、単に核酸の一部を付加、置換または欠失した変異核酸も含む。しばしば組換え核酸には、プロモーター配列に機能的に結合した核酸配列が含まれる。このような組換え核酸は、ベクターの不可欠なエレメントであって例えばある細胞を形質転換するために用いられるようなものであり得る。

【 0 0 8 0 】

或いはこのような組換え核酸は、ウイルスベクターの不可欠なエレメントであって例えばワクシニアウイルスに基づくものであり得る。ワクシニアウイルスは組換え核酸が発現する哺乳動物のワクチン接種に用いるもので、哺乳動物の防御免疫応答を誘導する。

10

【 0 0 8 1 】

「調節エレメント」は、通常は遺伝子の未翻訳領域に由来する核酸配列であり、エンハンサー、プロモーター、イントロン及び 5' 及び 3' の未翻訳領域（UTR）を含む。調節エレメントは、転写、翻訳または RNA 安定性を調節する宿主またはウイルスタンパク質と相互作用する。

【 0 0 8 2 】

「レポーター分子」とは、核酸、アミノ酸または抗体を標識するのに用いられる化学的または生化学的成分である。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、基質、補助因子、阻害因子、磁気粒子及びその他の当分野で既知の成分がある。

20

【 0 0 8 3 】

DNA 配列に関する「RNA 等価物」は、発生した窒素塩基チミンが全てウラシルに置換されていることと、糖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構成されていることを除いて、参照 DNA 配列と同一のヌクレオチド線形配列から構成されている。

【 0 0 8 4 】

「サンプル」の語は、その最も広い意味で用いられる。ATRS、ATRS をコードする核酸、またはその断片を含むと推定されるサンプルは、体液と、細胞からの抽出物や細胞から単離された染色体や細胞内小器官、膜と、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、組織プリント等を含み得る。

30

【 0 0 8 5 】

「特異結合」または「特異的に結合する」の語は、タンパク質またはペプチドと、アゴニスト、抗体、アンタゴニスト、小分子、任意の天然成分または合成結合成分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（例えば抗原決定基即ちエピトープ）であって結合分子が認識するものが存在するか否かに依存していることを意味している。例えば、抗体がエピトープ「A」に対して特異的である場合、標識された遊離した A 及びその抗体を含む反応において、エピトープ A（つまり遊離し、標識されていない A）を含むポリヌクレオチドの存在が、抗体に結合している標識された A の量を低減させる。

【 0 0 8 6 】

「実質上精製された」の語は、自然環境から取り除かれ、或いは単離または分離された核酸またはアミノ酸配列であって、自然に会合するその他の構成エレメントの少なくとも約 60%、好ましくは少なくとも約 75%、最も好ましいのは少なくとも約 90% が遊離しているものを指す。

40

【 0 0 8 7 】

「置換」は、1 若しくは数個のアミノ酸またはヌクレオチドを各々別のアミノ酸またはヌクレオチドに置換することを意味する。

【 0 0 8 8 】

「基質」は、任意の好適な固体または半固体の支持体を指すものであって、膜、フィルタ、チップ、スライド、ウェハ、ファイバー、磁性非磁性ビーズ、ゲル、管、プレート、ポリマー、微細粒子、毛管が含まれる。基質は、壁、溝、ピン、チャネル、孔等、様々な表

50

面形態を有することができ、基質表面にはポリヌクレオチドやポリペプチドが結合する。

【0089】

「転写イメージ」は、所与の時間、条件での固有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。

【0090】

「形質転換」は、外来性のDNAが宿主細胞に入り込み、宿主細胞を変化させるプロセスを表す。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、リポフェクション及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或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。さらに、限られた時間に一時的に導入DNA若しくは導入RNAを発現する細胞も含まれる。

10

【0091】

ここで用いる「遺伝形質転換体」とは任意の有機体であり、限定するものではないが動植物を含み、有機体の1個若しくは数個の細胞が、ヒトの関与によって、例えば本技術分野でよく知られている形質転換技術によって導入された異種核酸を有する。核酸の細胞への導入は、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、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或いは組換えウイルスの導入によって行う。遺伝子操作の語は、古典的な交雑育種或いは*in vitro*受精を指すものではなく、組換えDNA分子の導入を指すものである。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、真菌及び動植物がある。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をこのような有機体に移入する技術はよく知られており、前出Sambrook（1989）等の参考文献に記載されている。

20

【0092】

特定の核酸配列の「変異体」は、核酸配列1本全部の長さに対して特定の核酸配列と少なくとも40%の相同性を有する核酸配列であると定義する。その際、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%またはそれ以上の相同性を示し得る。或る変異体は、例えば「対立遺伝子」変異体（前述）、「スプライス」変異体、「種」変異体または「多形性」変異体として説明し得る。スプライス変異体は参照分子とかなりの相同性を有し得るが、mRNAプロセッシング中のエキソンの交互スプライシングによって通常多数の或いは僅かな数のポリ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、追加機能ドメインを有するか或いは参照分子に存在するドメインが欠落していることがある。種変異体は、種相互に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、相互にかなりのアミノ酸相同性を有する。多形性変異体は、与えられた種の個体間で特定の遺伝子のポリヌクレオチド配列が異なる。また、多形性変異体は、1つのヌクレオチド塩基によってポリヌクレオチド配列が変化する「単一ヌクレオチド多形性」（SNP）を含み得る。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。

30

40

【0093】

特定のポリペプチド配列の「変異体」は、ポリペプチド配列の1本の長さ全体で特定のポリペプチド配列に対して少なくとも40%の相同性を有するポリペプチド配列として画定される。ここで、デフォルトパラメータに設定した「BLAST 2 Sequence

50

s」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、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。

【0094】

(発明)

本発明は、新規のヒトアミノアシル tRNA 合成酵素 (ATRS) 及び ATRS をコードするポリヌクレオチドの発見に基づき、これらの組成物を利用した細胞増殖異常、自己免疫/炎症性の疾患の診断、治療、及び予防に関する。

【0095】

表1は、本発明の完全長ポリヌクレオチド配列及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及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) と相関する。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 ポリヌクレオチド配列番号 (Incyte ポリヌクレオチド ID) によって表示した。

【0096】

表2は、GenBank タンパク質 (genpept) データベースに対する BLAST 分析によって同定されたような、本発明のポリペプチドとの相同性を有する配列を示している。列1と2は発明したポリペプチドをポリペプチド配列識別番号 (ポリペプチド SEQ ID NO) と対応する 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。列3は、GenBank の最も近い相同体の GenBank の識別番号 (Genbank ID NO :) を示す。列4は、各ポリペプチドとその GenBank 相同体との間の一致を表す確率スコアを示す。列5は、GwnBank 相同体のアノテーションを示し、更に該当箇所には適当な引用文も示す。これらを引用することを以て本明細書の一部とする。

【0097】

表3は、本発明のポリペプチドの様々な構造的特徴を示す。列1と2は発明した個々のポリペプチドをポリペプチド配列識別番号 (SEQ ID NO) と対応する 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 (signature) 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、該当箇所にはさらに分析方法に利用した検索可能なデータベースを示す。

【0098】

表2及び表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アミノアシル tRNA 合成酵素であることを確立している。例えば、SEQ ID NO: 1 はアミノ酸残基 51 から 503 で、古細菌 (Pyrococcus abyssi) の システイニル tRNA 合成酵素 (GenBank ID g5458823) と 41% の同一性を有することが Basic Local Alignment Search Tool (BLAST) によって示された。(表2参照) BLAST 確率スコアは 1.4×10^{-81} 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、tRNA 合成酵素 クラス I (C) ドメインを有する。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) BLIMP S 解析から得られたデータによって、SEQ ID NO: 1 がシステイニル tRNA 合成酵素であ

10

20

30

40

50

ることが裏付けられた。別の例において、SEQ ID NO: 2 は Basic Local Alignment Search Tool (BLAST) によって同定される Synechocystis sp. アスパラギニル tRNA 合成酵素 (GenBank ID g1001357) に 46% の同一性を有する。(表 2 参照) BLAST 確率スコアは 7.8×10^{-4} 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: 2 はまた、tRNA 合成酵素 クラス II (D, K および N) ドメインを有する。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: 2 がアスパラギニル tRNA 合成酵素である、さらに実証的な証拠を提供する。別の例において、SEQ ID NO: 4 は Basic Local Alignment Search Tool (BLAST) によって同定される Bacillus caldotaenax チロシル tRNA 合成酵素 (GenBank ID g143793) に 39% の同一性を有する。(表 2 参照) BLAST 確率スコアは 2.3×10^{-72} 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: 4 はまた、チロシル tRNA 合成酵素 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: 4 が tRNA 合成酵素である、さらに実証的な証拠を提供する。SEQ ID NO: 3 は、同様の規則で解析され注釈された。SEQ ID NO: 1 - 4 の解析のためのアルゴリズム及びパラメータが表 7 で記述されている。

10

20

30

40

50

【0099】

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、或いはこれら 2 種類の配列を任意に組み合わせて構築した。列 1 と 2 は本発明のポリヌクレオチドをポリヌクレオチド配列識別番号 (Polynucleotide SEQ ID NO) と対応する 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) を示す。列 3 は、塩基対における各ポリヌクレオチド配列の長さを示す。列 4 は、例えば、SEQ ID NO: 5 - 8 を同定するため、或いは SEQ ID NO: 5 - 8 と関連するポリヌクレオチド配列とを区別するためのハイブリダイゼーションまたは増幅技術に有用なポリヌクレオチド配列の断片を示す。列 5 は cDNA 配列、ゲノム DNA から予想されたコード配列 (エキソン) 及び / または cDNA 及びゲノム DNA を共に有する配列集合に対応する識別番号を示している。これらの配列は、本発明の完全長ポリヌクレオチド配列をアセンブルするのに用いた。表 4 の列 6 および列 7 はそれぞれ、列 5 の配列に対応する cDNA 配列および / またはゲノム配列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。

【0100】

表 4 の列 5 の識別番号は、特に例えば Incyte cDNA とそれに対応する cDNA ライブラリに照会し得る。例えば、2700694F6 は Incyte cDNA 配列の識別番号であり、OVRTUT10 はそれが由来する cDNA ライブラリの識別番号である。または、列 5 の識別番号は、ポリヌクレオチド配列の組み立てに用いた GenBank の cDNA すなわち EST (例えば、g6451182) の識別番号の場合もある。cDNA ライブラリが示されていない Incyte cDNA は、プールされている cDNA ライブラリ (例えば、70997854V1) に由来する。或いは列 5 の識別番号は、ゲノム DNA の Genscan 分析により予想されるコード領域に照会し得る。この Genscan 推定コード配列は、配列を組み立てる前に編集する場合がある。(例 IV を参照)。または列 5 の識別番号は、「エキソンステッチング (exon-stitching)」アルゴリズムにより結び合わせた cDNA 及び Genscan 予想エキソンの両

方の集合に照会し得る。または、列5の識別番号は、「エキソンスティッチング」アルゴリズムによってcDNAおよびGenScan推定エキソンの両方からなる群の場合もある(実施例5を参照)。場合によっては、列5に示されている配列の範囲と重複するIncyte cDNAの範囲が得られ、最終的なコンセンサス配列が決定されるが、それに相当するIncyte cDNAの識別番号は示されていない。

【0101】

表5は、Incyte cDNA配列を用いてアセンブル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ライブラリは、上記のポリヌクレオチド配列をアセンブル及び確認するために用いられるIncyte cDNA配列によって最も頻繁に表されるIncyte cDNAライブラリである。cDNAライブラリを作製するために用いた組織及びベクターを表5に示し、表6で説明している。

10

【0102】

本発明はまた、ATRSの変異体も含む。好適なATRSの変異体は、ATRSの機能的或いは構造的特徴の少なくともどちらか一方を有し、かつATRS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。

【0103】

本発明はまた、ATRSをコードするポリヌクレオチドを提供する。特定の実施例において、本発明は、ATRSをコードするSEQ ID NO: 5-8からなる一群から選択された配列を含むポリヌクレオチド配列を提供する。SEQ ID NO: 5-8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配列と同等の価値を有しているが、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リボースから構成されている。

20

【0104】

本発明はまた、ATRSをコードするポリヌクレオチド配列の変異配列を含む。詳細には、このようなポリヌクレオチド配列の変異配列は、ATRSをコードするポリヌクレオチド配列と少なくとも80%のポリヌクレオチド配列同一性、或いは少なくとも95%のポリヌクレオチド配列同一性、更には少なくとも95%ものポリヌクレオチド配列同一性を有する。本発明の或る実施態様では、SEQ ID NO: 5-8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%もの一致率を有するようなSEQ ID NO: 5-8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配列を含む。上記の任意のポリヌクレオチドの変異体は、ATRS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。

30

【0105】

遺伝暗号の縮重により作り出され得るATRSをコードする種々のポリヌクレオチド配列には、既知の自然発生する任意の遺伝子のポリヌクレオチド配列と最小の類似性しか有しないものも含まれることを、当業者は理解するであろう。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ようなありとあらゆる可能性のあるポリヌクレオチド配列変異体を網羅し得る。これらの組み合わせは、天然のATRS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作られ、全ての変異が明確に開示されていると考慮する。

40

【0106】

ATRSをコードするヌクレオチド配列及びその変異配列は一般に、好適に選択されたストリンジェントな条件下で、天然のATRSのヌクレオチド配列とハイブリダイズ可能であるが、非天然のコドンを含めるなどの実質的に異なった使い方のコドンを含むATRS或いはその誘導体をコードするヌクレオチド配列を作るとは有利となり得る。宿主が特定のコドンを利用する頻度に基づいて、特定の真核又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することが可能である。コードされたアミノ酸配列を

50

変えないで、A T R S 及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変更する別の理由は、天然の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転写物を作ることにある。

【0107】

本発明はまた、A T R S 及びその誘導体をコードするDNA配列またはそれらの断片を完全に合成化学によって作り出すことも含む。作製後、当分野でよく知られている試薬を用いて、この合成配列を任意の様々な入手可能な発現ベクター及び細胞系中に挿入し得る。

更に、合成化学を用いてA T R S またはその任意の断片をコードする配列に突然変異を誘導し得る。

【0108】

さらに、本発明には、種々のストリンジェントな条件下で、請求項に記載されたポリヌクレオチド配列、特に、SEQ ID NO: 5 - 8 及びそれらの断片とハイブリダイズ可能なポリヌクレオチド配列が含まれる（例えば、Wahl, G. M. 及び S. L. Berger (1987) *Methods Enzymol.* 152: 399 - 407; Kimmel, A. R. (1987) *Methods Enzymol.* 152: 507 - 511. を参照）。アニーリング及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」に記載されている。

【0109】

DNAシーケンシングの方法は当分野でよく知られており、本発明の何れの実施例もDNAシーケンシング方法を用いて実施可能である。DNAシーケンシング方法には酵素を用いることができ、例えばDNAポリメラーゼI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) を用いることができる。或いは、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を併用することができる。好適には、MICROLAB 2200液体転移システム (Hamilton, Reno, NV)、PTC 200サーマルサイクラー (MJ Research, Watertown MA) 及びABI CATALYST 800サーマルサイクラー (Applied Biosystems) 等の装置を用いて配列の準備を自動化する。次に、ABI 373 或いは 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) または当分野でよく知られている他の方法を用いてシーケンシングを行う。結果として得られた配列を当分野でよく知られている種々のアルゴリズムを用いて分析する。 (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, 856 - 853ページ等を参照)。

【0110】

当分野で周知のPCR法をベースにした種々の方法で、部分的なヌクレオチド配列を利用して、A T R S 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の上流にある配列を検出する。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及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から未知の配列を増幅する方法である (Sarkar, G. (1993) *PCR Methods Applic.* 2: 318322 を参照)。別の方法に逆PCR法があり、これは広範な方向に伸長させたプライマー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、既知のゲノム遺伝子座及びその周辺の配列からなる制限酵素断片から得る (Triglia, T. 他 (1988) *Nucleic Ac*

10

20

30

40

50

ids Res. 16:8186. を参照)。第3の方法としてキャプチャPCR法があり、これはヒト及び酵母菌人工染色体DNAの既知の配列に隣接するDNA断片をPCR増幅する方法に關与している。(Lagerstrom, M. 他(1991) PCR Methods Applic. 1:111119. 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及び連結反応を用いて未知の配列領域内に組換え二本鎖配列を挿入することが可能である。また、未知の配列を検索するために用い得る別の方法については当分野で知られている。(Parker, J. D. ら(1991) Nucleic Acids Res. 19:30553060 を参照)。更に、PCR、ネステッドプライマー及びPromoter Finder(商標)ライブラリ(Clontech, Palo Alto CA)を用いてゲノムDNAをウォーキングすることが

10

できる。この手順は、ライブラリをスクリーニングする必要がなく、イントロン/エキソン接合部を見付けるのに有用である。全てのPCRベースの方法に対して、市販されているソフトウェア、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、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を設計し得る。

【0111】

完全長cDNAをスクリーニングする際は、より大きなcDNAを含むようにサイズ選択されたライブラリ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は、しばしば遺伝子の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。ゲノムライブラリは、5'非転写調節領域への配列の伸長に有用であろう。

20

【0112】

市販されているキャピラリー電気泳動システムを用いて、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することができる。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、レーザで活性化される蛍光色素と、放出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等)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及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。

30

【0113】

本発明の別の実施例では、ATRSをコードするポリヌクレオチド配列またはその断片を組換えDNA分子にクローニングして、適切な宿主細胞内にATRS、その断片または機能的等価物を発現させることが可能である。遺伝暗号固有の縮重により、実質的に同じ或いは機能的に等価のアミノ酸配列をコードする別のDNA配列が作られ得り、これらの配列をATRSのクローン化及び発現に利用可能である。

【0114】

種々の目的でATRSをコードする配列を変えるために、当分野で一般的に知られている方法を用いて、本発明のヌクレオチド配列を組換えることができる。この目的には、遺伝子産物のクローン化、プロセッシング及び/または発現の調節が含まれるが、これらに限定されるものではない。遺伝子断片及び合成オリゴヌクレオチドのランダムなフラグメンテーション及び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み換えることが可能である。例えば、オリゴヌクレオチド仲介定方向突然変異誘導を利用して、新規な制限部位の生成、グリコシル化パターンの変更、コドン優先の変更、スプライス変異体の生成等を行う突然変異を導入し得る。

【0115】

本発明のヌクレオチドは、MOLECULAR BREEDING (Maxygen In

40

50

c., 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もので、生物学的または酵素的な活性、或いは他の分子や化合物と結合する能力などのATRSの生物学的特性を変更或いは改良することができる。DNAシャッフリングは、遺伝子断片のPCR伸介再組換えを用いて遺伝子変異体のライブラリを生成するプロセスである。ライブラリはその後、その遺伝子変異体を所望の特性に同定するような選択またはスクリーニングにかける。次にこれらの好適な変異体をプールし、更に反復してDNAシャッフリング及び選択/スクリーニングを行ってもよい。このように、遺伝の多様性は「人為的」品種改良及び急速な分子の進化を経て創生される。例えば、ランダムポイント突然変異を有する単一の遺伝子の断片を組み換えて、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することができる。或いは、所定の遺伝子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と組み換え、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、指図された制御可能な方法で最大化させることができる。

10

【0116】

別の実施例によれば、ATRSをコードする配列は、当分野で周知の化学的方法を用いて、全体或いは一部が合成可能である(例えば、Caruthers, M.H.ら(1980)(1980) Nucl. Acids Res. Symp. Ser 7:215-223; 及びHorn, T.他(1980) Nucl. Acids Res. Symp. Ser. 7:225-232を参照)。別法として、化学的方法を用いてATRS自体またはその断片を合成することが可能である。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うことができる(Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, 55-60ページ、Roberge, J.Y. ら(1995) Science 269:202-204等を参照)。自動合成はABI 431Aペプチドシンセサイザ(Perkin Elmer)を用いて達成し得る。更にATRSのアミノ酸配列または任意のその一部は、直接的な合成の際の変更、及び/または化学的方法を用いた他のタンパク質または任意のその一部からの配列との組み合わせにより、天然のポリペプチド配列を有するポリペプチドまたは変異体ポリペプチドを作製することが可能である。

20

30

【0117】

ペプチドは、分離用高速液体クロマトグラフィーを用いて実質上精製可能である(Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421等を参照)。(Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421等を参照)。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することができる(前出のCreighton, 28-53ページ等を参照)。

40

【0118】

生物学的に活性なATRSを発現させるために、ATRS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する。この発現ベクターは、好適な宿主に挿入されたコーディング配列の転写及び翻訳の調節に必要なエレメントを含む。これらのエレメントには、ベクター及びPPをコードするポリヌクレオチド配列におけるエンハンサー、構成型及び発現誘導型のプロモーター、5'及び3'の非翻訳領域などの調節配列が含まれる。このような要素は、長さ及び特異性が様々である。特定の開始シグナルによって、ATRSをコードする配列のより効果的な翻訳を達成することが可能である。このようなシグナルには、ATG開始コドンと、コザック配列などの近傍の配列が含まれる

50

。A T R Sをコードする配列及びその開始コドン、上流の調節配列が好適な発現ベクターに挿入された場合は、更なる転写調節シグナルや翻訳調節シグナルは必要なくなるであろう。しかしながら、コーディング配列或いはその断片のみが挿入された場合は、インフレームのA T G開始コドンを含む外来性の翻訳調節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。外来性の翻訳要素及び開始コドンは、様々な天然物及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 (Scharf, D. 他 (1994) Results Probl. 20:125162を参照)。

【0119】

当業者に周知の方法を用いて、A T R Sをコードする配列、好適な転写及び翻訳調節エレメントを含む発現ベクターを作製することが可能である。これらの方法には、in vitro組換えDNA技術、合成技術、及びin vivo遺伝子組換え技術が含まれる (例えば、Sambrook, J. 他. (1989) Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, 4章及び8章, 及び16-17章; 及び Ausubel, F.M. 他. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, 9章、13章及び16章を参照)。

【0120】

種々の発現ベクター/宿主系を利用して、A T R Sをコードする配列の保持及び発現が可能である。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、動物細胞系などの微生物等がある。(前出のSambrook、前出のAusubel、Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509、Engelhard, E.K. ら (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227、Sandig, V. ら (1996) Hum. 7:1937-1945、Takamatsu, N. (1987) EMBOJ. 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. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659、Harrington, J.J. ら (1997) Nat. Genet. 15:345-355等を参照)。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輸送することができる (Di Nicola, M. ら (1998) Cancer Gen. Ther. 5(6):350-356、Yu, M. ら (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344、Buller, R.M. ら (1985) Nature 317(6040):813-815; McGregor, D.P. ら (1994) Mol. Immunol. 31(3):219-226、Verma, I.M. and N. Somia (1997) Nature 389:239-242等を参照)。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。

【0121】

細菌系では、多数のクローニングベクター及び発現ベクターが、A T R S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選択可能である。例えば、A T R Sをコードするポリヌクレオチド配列の日常的なクローニング、サブクローニング、増殖には、P B L U E S C R I P T (S t r a t a g e n e , L a J o l l a C A) または p S P O R T 1 プラスミド (G I B C O B R L) などの多機能の大腸菌ベクターを用いることができる。ベクターの多数のクローニング部位にA T R Sをコードする配列をライゲーションするとl a c Z 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i n v i t r o 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救出、入れ子状態の欠失の生成にも有用であろう(例えば、V a n H e e k e , G . および S . M . S c h u s t e r (1 9 8 9) J . B i o l . C h e m . 2 6 4 : 5 5 0 3 5 5 0 9 を参照)。例えば、抗体の産生のためなどに多量のA T R Sが必要な場合は、A T R S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T 5 バクテリオファージプロモーターまたは誘導T 7 バクテリオファージプロモーターを含むベクターが使用できる。

10

【0122】

A T R Sの発現に酵母の発現系の使用が可能である。因子、アルコールオキシダーゼ、P G Hプロモーター等の構成型或いは誘導型のプロモーターを含む多数のベクターが、酵母菌サッカロミセス・セレビジエまたは*P i c h i a p a s t o r i s*に使用可能である。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への保持のどちらかを誘導し、安定した増殖のために宿主ゲノムの中に外来配列を組み込む。(例えば、A u s u b e l , 1 9 9 5 , 前出、B i t t e r , G . A . ら (1 9 8 7) M e t h o d s E n z y m o l . 1 5 3 : 5 1 6 - 5 4 4、及びS c o r e r , C . A . ら (1 9 9 4) B i o / T e c h n o l o g y 1 2 1 - 1 8 1 - 1 8 4 . を参照)。

20

【0123】

植物系もA T R Sの発現に使用可能である。A T R S をコードする配列の転写は、ウイルスプロモーター、例えば単独或いはT M V (T a k a m a t s u , N . (1 9 8 7) E M B O J 6 : 3 0 7 - 3 1 1) 由来のオメガリーダー配列と組み合わせ用いられるようなC a M V 由来の3 5 S 及び1 9 S プロモーターによって促進される。或いは、R U B I S C O の小サブユニット等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てもよい(前出のC o r u z z i (1 9 8 4)、E M B O J . 3 : 1 6 7 1 1 6 8 0、前出のB r o g l i e (1 9 8 4) S c i e n c e 2 2 4 : 8 3 8 8 4 3、前出のW i n t e r (1 9 9 1) R e s u l t s P r o b l . C e l l D i f f e r . 1 7 : 8 5 1 0 5 等を参照)。これらの構成物は、直接DNA形質転換または病原体を媒介とする形質移入によって、植物細胞内に導入可能である。(『マグローヒル科学技術年鑑』(The McGraw Hill Yearbook of Science and Technology) (1 9 9 2) M c G r a w H i l l N e w Y o r k N Y , 1 9 1 - 1 9 6 ページ等を参照)。

30

【0124】

哺乳動物細胞においては、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、後発プロモーター及び3 連リーダー配列からなるアデノウイルス転写物/翻訳複合体にA T R Sをコードする配列を結合し得る。可欠E 1 またはE 3 領域へウイルスのゲノムを挿入し、宿主細胞でA T R Sを発現する感染ウイルスを得ることが可能である。(例えば、L o g a n , J . および T . S h e n k (1 9 8 4) P r o c . N a t l . A c a d . S c i . U S A 8 1 : 3 6 5 5 3 6 5 9 を参照)。更に、ラウス肉腫ウイルス(R S V) エンハンサー等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。S V 4 0 またはE B V 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。

40

【0125】

ヒト人工染色体(H A C)を用いて、プラスミドに含まれ且つプラスミドから発現するものより大きなDNAの断片を輸送することもできる。治療のために約6 k b ~ 1 0 M b の

50

H A C s を作製し、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）で供給する。（例えば、Harrington, J. J. 他（1997）Nat Genet. 15: 345 - 355. を参照）。

【0126】

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は、株化細胞における A T R S の安定した発現が望ましい。例えば、発現ベクターを用いて、A T R S をコードする配列を株化細胞に形質転換することが可能である。このような発現ベクターは、ウイルス起源の複製及び／または内在性の発現要素や、同じ或いは別のベクターの上の選択マーカー遺伝子を含む。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ることができる。選択可能マーカーの目的は選択培地への抵抗性を与えることであり、選択可能マーカーが存在することにより、導入された配列をうまく発現するような細胞の成長及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

【0127】

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、t k⁺ 単純細胞のために用いられるヘルペスウイルスチミジンキナーゼ遺伝子と、a p r⁺ 細胞のために用いられる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（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、抗生物質或いは除草剤への耐性を用いることができる。例えば d h f r はメトトレキセートに対する耐性を与え、n e o はアミノグリコシッドネオマイシン及び G - 4 1 8 に対する耐性を与え、a l s はクロルスルフロンに対する耐性を、p a t 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; ColbereGarapin, F. 他（1981）J. Mol. Biol. 150: 114 を参照）。この他の選択可能な遺伝子、例えば、代謝のための細胞の必要条件を変える t r p B 及び h i s D は、文献に記載されている（Hartman, S. C. および R. C. Mulligan（1988）Proc. Natl. Acad. Sci. USA 85: 8047 - 8051 等を参照）。可視マーカー、例えばアントシアニン、緑色蛍光タンパク質（G F P ; Clontech）、グルクロニダーゼ及びその基質 グルクロニド、またはルシフェラーゼ及びその基質ルシフェリン等を用いてもよい。これらのマーカーを用いて、トランスフォーマントを特定するだけでなく、特定のベクター系に起因する一過性或いは安定したタンパク質発現を定量することが可能である（Rhodes, C. A.（1995）Methods Mol. Biol. 55: 121131 等を参照）。

20

30

【0128】

マーカー遺伝子発現の存在／不存在によって目的の遺伝子の存在が示唆されても、その遺伝子の存在及び発現の確認が必要な場合もある。例えば、A T R S をコードする配列がマーカー遺伝子配列の中に挿入された場合、A T R S をコードする配列を含む形質転換された細胞は、マーカー遺伝子機能の欠落により特定可能である。または、1つのプロモーターの制御下でマーカー遺伝子が A T R S をコードする配列と一列に配置することも可能である。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。

40

【0129】

一般に、A T R S をコードする核酸配列を含み且つ A T R S を発現する宿主細胞は、当業者によく知られている種々の方法を用いて同定することが可能である。限定するものではないが当業者によく知られている方法には、D N A - D N A 或いは D N A - R N A ハイブリダイゼーション、P C R 法、核酸或いはタンパク質の検出、定量、或いはその両方を行うための膜系、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の生物学的検定法または免疫学的検定法がある。

50

【0130】

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 A T R S の発現の検出及び計測のための免疫学的な方法は、当分野で周知である。このような技法には、酵素に結合した免疫吸着剤検定法 (E L I S A)、ラジオイムノアッセイ (R I A)、フローサイトメーター (F A C S) などがある。A T R S 上の 2 つの非干渉エピトープに反応するモノクローナル抗体を用いた、2 部位のモノクローナルベース イムノアッセイ (two-site, monoclonal-based immunoassay) が好ましいが、競合の結合アッセイも用いることもできる。これらのアッセイ及びこれ以外のアッセイは、当分野で公知である (Hampton, R. ら (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN, Sect. IV, Coligan, J. E. ら (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY, Pound, J. D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ 等を参照)。

10

【0131】

多岐にわたる標識方法及び結合方法が、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。A T R S 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 P C R 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、末端標識化、または標識されたヌクレオチドを用いる P C R 増幅が含まれる。別法として、A T R S をコードする配列、またはその任意の断片を m R N A 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、T 7、T 3 または S P 6 等の好適な R N A ポリメラーゼ及び標識されたヌクレオチドを加えて、in vitro で R N A 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 等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或いは標識には、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、発色剤等がある。

20

30

【0132】

A T R S 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及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、或いはその両者に依存する。A T R S をコードするポリヌクレオチドを含む発現ベクターは、原核細胞膜及び真核細胞膜を透過する A T R S の分泌を誘導するシグナル配列を含むように設計できることは、当業者には理解されよう。

【0133】

更に、宿主細胞株の選択は、挿入した配列の発現を調節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、脂質化及びアシル化がある。タンパク質の「プレプロ」または「プロ」形を切断するような翻訳後処理を利用して、タンパク質のターゲティング、折りたたみ及び/または活性を特定することも可能である。翻訳後の活性のための固有の細胞装置及び特徴のある機構を有する種々の宿主細胞 (例えば CHO、HeLa、MDCK、MEK 293、WI38 等) は、American Type Culture Collection (ATCC, Bethesda, VA) から入手可能であり、外来タンパク質の正しい修飾及び処理を確実にするように選択し得る。

40

【0134】

50

本発明の別の実施例では、A T R Sをコードする自然或いは変更された、または組換えの核酸配列を上記した任意の宿主系の融合タンパク質の翻訳となる異種配列に結合させる。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A T R Sタンパク質が、A T R S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及び異種ペプチド部分も、市販されている親和性基質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 (G S T)、マルトース結合タンパク質 (M B P)、チオレドキシン (T r x)、カルモジュリン結合ペプチド (C B P)、6 - H i s、F L A G、c - m y c、赤血球凝集素 (H A) がある。G S Tは固定化グルタチオン上で、M B Pはマルトース上で、T r xはフェニルアルシンオキシド上で、C B Pはカルモジュリン 10
上で、そして6 - H i s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。F L A G、c - m y c及び赤血球凝集素 (H A) は、これらのエピトープ標識を特異的に認識する市販されているモノクローナル抗体及びポリクローナル抗体を用いて、融合タンパク質の免疫親和性精製を可能にする。また、A T R Sをコードする配列と異種タンパク質配列との間にあるタンパク質分解切断部位を融合タンパク質が含むように遺伝子操作すると、A T R Sが精製の後に異種部分から切断され得る。融合タンパク質の発現及び精製方法は、前出のA u s u b e l (1 9 9 5) 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及び精製を促進することもできる。

【0135】

本発明の別の実施例では、T N Tウサギ網状赤血球可溶化液またはコムギ胚芽抽出系 (P r o m e g a) を用いて i n v i t r o で放射能標識したA T R Sの合成が可能である 20
。これらの系は、T 7、T 3またはS P 6プロモーターと機能的に結合したタンパク質コード配列の転写及び翻訳を結合する。翻訳は、例えば³ S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。

【0136】

本発明のA T R Sまたはその断片を用いて、A T R Sに特異結合する化合物をスクリーニングすることができる。少なくとも1つまたは複数の試験化合物を用いて、A T R Sへの特異的な結合をスクリーニングすることが可能である。試験化合物の例には、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。

【0137】

一実施例では、このように同定された化合物は、例えばリガンドやその断片などのA T R Sの天然のリガンド、または天然の基質、構造的または機能的な擬態性または自然結合パートナーに密接に関連している (C o l l i g a n , J . E . 他 (1 9 9 1) C u r r e n t P r o t o c o l s i n I m m u n o l o g y 1 (2) の5章等を参照)。同様に、化合物は、A T R Sが結合する天然受容体、或いは例えばリガンド結合部位などの少なくとも受容体のあ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。一実施例では、このような化合物に対するスクリーニングには、分泌タンパク質或いは細胞膜上のタンパク質の何れか一方としてA T R Sを発現する好適な細胞の作製が含まれる。好適な細胞には、哺乳動物、酵母、大腸菌からの細胞が含まれる。A T R Sを発現する細胞またはA T R Sを含有する細胞膜断片を 40
試験化合物と接触させて、A T R Sまたは化合物の何れかの結合、刺激または阻害を分析する。

【0138】

あるアッセイは、単に試験化合物をポリペプチドに実験的に結合させ、結合を、蛍光色素、放射性同位体、酵素抱合体またはその他の検出可能な標識により検出することができる。例えば、このアッセイは、少なくとも1つの試験化合物を、溶液中の或いは固体支持物に固定されたA T R Sと結合させるステップと、A T R S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及び測定を行うことができる。更にこのアッセイでは、無細胞再構成標本、化学ライブラリまたは天然の生成混合物を用いて実施することができ、試験化合物は、溶液中で遊離 50

させるか固体支持体に固定させる。

【0139】

本発明のA T R Sまたはその断片を用いて、A T R Sの活性を調整する化合物をスクリーニングすることが可能である。このような化合物には、アゴニスト、アンタゴニスト、或るいは部分的または逆アゴニスト等が含まれる。一実施例では、A T R Sが少なくとも1つの試験化合物と結合する、A T R Sの活性が許容される条件下でアッセイを実施し、試験化合物の存在下でのA T R Sの活性が試験化合物不在下でのA T R Sの活性と比較する。試験化合物の存在下でのA T R Sの活性の変化は、A T R Sの活性を調整する化合物の存在を示唆する。別法では、試験化合物をA T R Sの活性に適した条件下でA T R Sを含む i n v i t r o または無細胞再構成系と結合させてアッセイを実施する。これらアッセイの何れかにおいて、A T R Sの活性を調整する試験化合物は間接的に結合することが可能であり、試験化合物と直接接触する必要がない。少なくとも1つから複数の試験化合物をスクリーニングすることができる。

10

【0140】

別の実施例では、胚性幹細胞（E S細胞）における相同組換えを用いて動物モデル系内で、A T R S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「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号及び第5,767,337号等を参照）。例えば129/SvJ細胞株等のマウスE S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E S細胞は、ネオマイシンホスホトランスフェラーゼ遺伝子（n e o : C a p e c c h i , M . R . (1 9 8 9) S c i e n c e 2 4 4 : 1 2 8 8 - 1 2 9 2) 等のマーカー遺伝子で破壊した目的の遺伝子を含む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み込まれる。別法では、C r e - l o x P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（M a r t h , J . D . (1 9 9 6) C l i n . I n v e s t . 9 7 : 1 9 9 9 - 2 0 0 2 ; W a g n e r , K . U . 他 (1 9 9 7) N u c l e i c A c i d s R e s . 2 5 : 4 3 2 3 - 4 3 3 0) 。形質転換したE S細胞を同定し、例えばC57BL/6マウス系等から採取したマウス細胞胚盤胞に微量注入する。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決め、これ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 30 遺伝子組換え動物は、可能性のある治療薬や毒性薬剤で検査することができる。

20

30

【0141】

A T R Sをコードするポリヌクレオチドを i n v i t r o でヒト胚盤胞由来のE S細胞において操作することが可能である。ヒトE S細胞は、内胚葉、中胚葉及び外胚葉の細胞の種類を含む少なくとも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、造血系統及び心筋細胞に分化する（T h o m s o n , J . A . 他 (1 9 9 8) S c i e n c e 2 8 2 : 1 1 4 5 - 1 1 4 7) 。

【0142】

A T R S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）を作製することが可能である。ノックイン技術を用いて、A T R Sをコードするポリヌクレオチドの或る領域を動物E S細胞に注入し、注入した配列を動物細胞ゲノムに組み込ませる。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研究し、可能性のある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。別法では、例えばA T R Sを乳汁内に分泌するなどA T R Sを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（J a n n e , J . 他 (1 9 9 8) B i o t e c h n o l . A n n u . R e v . 4 : 5 5 - 7 4 を参照）。

40

【0143】

（治療）

A T R Sのある領域とアミノアシル t R N A 合成酵素の領域との間に、例えば配列及び

50

モチーフの内容における化学的及び構造的類似性が存在する。更に、A T R S の発現は、肺腫瘍組織、新生児ケラチン生成細胞、リンパ腺細胞、結腸および前立腺の疾患状態に密接に関連する。従って、A T R S は、細胞増殖異常、自己免疫 / 炎症性の疾患においてある役割を果たすと考えられる。A T R S 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、A T R S の発現または活性を低下させることが望ましい。また、A T R S 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、A T R S の発現または活性を増大させることが望ましい。

【 0 1 4 4 】

従って、一実施例において、A T R S 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、患者にA T R S またはその断片や誘導体を投与することが可能である。限定するものではないがこのような疾患の例として細胞増殖異常も含まれ、その中には日光性角化症、動脈硬化症、アテローム性動脈硬化症、滑液包炎、硬変、肝炎、混合型結合組織病 (M C T D)、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、原発性血小板血症と、腺癌、白血病、リンパ腫、黒色腫、骨髓腫、肉腫、奇形癌を含む癌、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、子宮の癌等が含まれる。また自己免疫 / 炎症疾患も含まれ、その中には後天性免疫不全症候群 (A I D S)、アジソン病、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症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫多発性内分泌腺障害症 (A P E C E D)、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、リンパ球毒素性一時的リンパ球減少症、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、過敏性大腸症候群、多発性硬化症、重症筋無力症、心筋または心膜の炎症、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、全身性紅斑性狼瘡、全身性硬化症、血小板減少症、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌の合併症、血液透析、体外循環、ウイルス性感染症、細菌性感染症、真菌性感染症、寄生虫感染症、原虫感染症、蠕虫の感染症及び外傷が含まれる。

【 0 1 4 5 】

別の実施例では、限定するものではないが上に列記した疾患を含むA T R S 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、A T R S またはその断片や誘導体を発現し得るベクターを患者に投与することも可能である。

【 0 1 4 6 】

更に別の実施例では、限定するものではないが上に列記した疾患を含むA T R S 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A T R S を含む組成物を好適な医薬用担体と共に患者に投与することも可能である。

【 0 1 4 7 】

更に別の実施例では、限定するものではないが上に列記した疾患を含むA T R S 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、A T R S の活性を調節するアゴニストを患者に投与することも可能である。

【 0 1 4 8 】

更なる実施例では、A T R S 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、患者にA T R S のアンタゴニストを投与することが可能である。限定するものではないが、このような疾患の例には、上記した細胞増殖異常、自己免疫 / 炎症性の疾患が含まれる。一実施態様では、A T R S と特異的に結合する抗体が直接アンタゴニストとして、或いはA T R S 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或いは送達機構として間接的に用いられ得る。

【 0 1 4 9 】

別の実施例では、限定するものではないが上に列記した疾患を含むA T R S 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、A T R S 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患者に投与することも可能である。

【0150】

別の実施例では、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み合わ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を選択し得る。治療薬と組み合わせることにより、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。この方法を用いることにより少量の各薬剤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。

【0151】

A T R S のアンタゴニストは、当分野で一般的な方法を用いて製造することが可能である。詳しくは、精製された A T R S を用いて抗体を作ったり、治療薬のライブラリをスクリーニングして A T R S と特異的に結合するものの同定が可能である。A T R S の抗体も、当分野で一般的な方法を用いて製造することが可能である。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、F a b 断片、及び F a b 発現ライブラリによって作られた断片が含まれる。但し、これらに限定されるものではない。中和抗体（即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。

【0152】

抗体の産生のためには、ヤギ、ウサギ、ラット、マウス、ヒト及びその他のものを含む種々の宿主が、A T R S または任意の断片、または免疫原性の特性を備えるそのオリゴペプチドの注入によって免疫化され得る。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等のミネラルゲルアジュバントと、リゾレシチン、ブルロニックポリオール、ポリアニオン、ペプチド、油性乳剤、キーホールリンペットヘモシニアン、ジニトロフェノール等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、B C G（カルメット ゲラン杆菌）及びコリネバクテリウム パルヴムが特に好ましい。

【0153】

A T R S に対する抗体を誘発するために用いられるオリゴペプチド、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り、一般的には約 10 個以上のアミノ酸からなるものが好ましい。これらのオリゴペプチド、ペプチドまたは断片は、天然のタンパク質のアミノ酸配列の一部と同一であり且つ小さな天然の分子の全アミノ酸配列を含むことが望ましい。A T R S アミノ酸の短いストレッチは、K L H などの別のタンパク質の配列と融合し、キメラ分子に対する抗体が産生され得る。

【0154】

A T R S 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、抗体分子を産生する任意の技術を用いて作製することが可能である。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及び E B V - ハイブリドーマ技術がある（K o h l e r , G . ら（1975）Nature 256 : 495 - 497、K o z b o r , D . ら（1985）J . Immunol . Methods 81 : 31 - 42、C o t e , R . J . ら（1983）Proc . Natl . Acad . Sci . USA 80 : 2026 - 2030、C o l e , S . P . ら（1984）Mol . Cell Biol . 62 : 109 - 120等を参照）。

【0155】

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、好適な抗原特異性及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、M o r r i s o n , S . L . 他（1984）Proc . Natl . Acad . Sci . 81 : 6851 - 6855；N e u b e r g e r , M . S . 他（1984）Nature 312 : 604 - 608；T a k e d a , S . ら（1985）Nature 314 : 452 , 454を参照）。別法では、当分野

10

20

30

40

50

で周知の方法を用いて、一本鎖抗体の産生のための記載された技術を適用して、A T R S 特異的な一本鎖抗体を生成する。関連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 (B u r t o n D . R . (1 9 9 1) P r o c . N a t l . A c a d . S c i . U S A 8 8 : 1 0 1 3 4 - 1 0 1 3 7 等を参照)。

【0156】

抗体の産生は、リンパ球集団における *i n v i v o* 産生の誘導によって、或いは免疫グロブリンライブラリのスクリーニングまたは文献に開示されているような高特異結合試薬のパネルのスクリーニングによっても行い得る (O r l a n d i , R . ら (1 9 8 9) P r o c . N a t l . A c a d . S c i . U S A 8 6 : 3 8 3 3 - 3 8 3 7 、 W i n t e r , G . ら (1 9 9 1) N a t u r e 3 4 9 : 2 9 3 - 2 9 9 等を参照)。

10

【0157】

A T R S に対する特異的な結合部位を含む抗体も得ることができる。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 $F(a b')_2$ 断片と、 $F(a b')_2$ 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 $F a b$ 断片とがある。或いは、 $F a b$ 発現ライブラリを作製することによって、モノクローナル $F a b$ 断片を所望の特異性と迅速且つ容易に同定することが可能となる (H u s e , W . D . ら (1 9 8 9) S c i e n c e 2 5 6 : 1 2 7 5 - 1 2 8 1 等を参照)。

20

【0158】

種々のイムノアッセイを用いてスクリーニングし、所望の特異性を有する抗体を同定することができる。隔離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競合的な結合、または免疫放射線活性のための数々のプロトコルが、当分野では周知である。通常このようなイムノアッセイには、A T R S とその特異性抗体との間の複合体形成の計測が含まれる。二つの非干渉性 A T R S エピトープに対して反応性のモノクローナル抗体を用いる、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、競合的結合アッセイも利用することができる (P o u n d 、前出)。

【0159】

ラジオイムノアッセイ技術と共に S c a t c h a r d 分析などの様々な方法を用いて、A T R S に対する抗体の親和性を評価する。親和性を結合定数 $K a$ で表すが、この $K a$ は、平衡状態の下で A T R S 抗体複合体のモル濃度を遊離抗体と遊離抗原のモル濃度で除して得られる値である。多数の A T R S エピトープに対して親和性が不均一なポリクローナル抗体試薬の $K a$ は、A T R S に対する抗体の平均親和性または結合活性を表す。特定の A T R S エピトープに単一特異的なモノクローナル抗体試薬の $K a$ は、親和性の真の測定値を表す。 $K a$ 値が $10^9 \sim 10^{12}$ l i t e r / m o l の高親和性抗体試薬は、A T R S 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用いるのが好ましい。 $K a$ 値が $10^6 \sim 10^7$ l i t e r / m o l の低親和性抗体試薬は、A T R S が抗体から最終的に活性化状態で解離する必要がある免疫精製 (i m m u n o p u r i f i c a t i o n) 及び類似の処理に用いるのが好ましい (C a t t y , D . (1 9 8 8) A n t i b o d i e s , V o l u m e I : A P r a c t i c a l A p p r o a c h . I R L P r e s s , W a s h i n g t o n , D C ; L i d d e l l , J . E . 及び C r y e r , A . (1 9 9 1) A P r a c t i c a l G u i d e t o M o n o c l o n a l A n t i b o d i e s , J o h n W i l e y & S o n s , N e w Y o r k N Y) 。

30

40

【0160】

ポリクローナル抗体試薬の抗体価及び結合活性を更に評価して、或る下流の適用例に対するこのような試薬の品質及び適性を決定することができる。例えば、少なくとも $1 \sim 2$ m g / m l の特異的な抗体、好ましくは $5 \sim 10$ m g / m l の特異的な抗体を含むポリクローナル抗体試薬は一般に、A T R S 抗体複合体を沈殿させなければならない処理に用いら

50

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。(前出のCattlyの文献、同Colliganらの文献等を参照)。

【0161】

本発明の別の実施例では、ATRS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、治療目的で使うことができる。ある実施態様では、ATRS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及びRNA、修飾ヌクレオチド)を設計して遺伝子発現を変更することができる。このような技術は当分野では周知であり、アンチセンスオリゴヌクレオチドまたは大きな断片が、ATRSをコードする配列の制御領域から、またはコード領域に沿ったさまざまな位置から設計可能である。(Agrawal, S. 他, ed. (1996) Antisense Therapeutics, Humana Press Inc., Totowa NJを参照)。

10

【0162】

治療に用いる場合、アンチセンス配列を好適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発現する発現プラスミドの形で細胞内に送達することが可能である(Slater, J. E. ら (1998) J. Allergy Clin. Immunol. 102(3):469-475 及び Scanlon, K. J. ら (1995) 9(13):1288-1296. 等を参照)。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関連ウイルスベクター等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(Miller, A. D. (1990) Blood 76:271、前出のAusubel、Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347等を参照)。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及び当分野で公知のその他の系が含まれる(Rossi, J. J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R. J. ら (1998) J. Pharm. Sci. 87(11):1308-1315、Morris, M. C. ら (1997) Nucleic Acids Res. 25(14):2730-2736. 等を参照)。

20

30

【0163】

本発明の別の実施例では、ATRSをコードするポリヌクレオチドを、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。遺伝子治療を行うことにより、(i) 遺伝子欠損症(例えばX染色体鎖遺伝(Cavazzana-Calvo, M. 他 (2000) Science 288:669-672)により特徴付けられる重度の複合型免疫欠損(SCID)-X1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度の複合型免疫欠損(Blaese, R. M. 他 (1995) Science 270:475-480、Bordignon, C. 他 (1995) Science 270:470-475)、嚢胞性繊維症(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(thalassemia)、家族性高コレステロール血症、第VIII因子若しくは第IX因子欠損に起因する血友病(Crystal, R. G. (1995) Science 270:404-410、Verma, I. M. and Somia, N. (1997) Nature 389:239-242)を治療し、(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、Poeschl, E. 他 (1996) Proc. Natl. Acad. Sci. USA. 93:11395-113

40

50

99)、B型若しくはC型肝炎ウイルス(HBV、HCV)、Candida albicans及びParacoccidioides brasiliensis等の真菌寄生菌、並びにPlasmodium falciparum及びTrypanosoma cruzi等の原生動物寄生体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。ATRS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ATRS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。

【0164】

本発明の更なる実施例では、ATRSの欠損による疾患や異常症は、ATRS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ATRS欠損細胞に導入することによって治療する。in vivo或いは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)弾道的金粒子の打ち込み(ballistic gold particle delivery)、(iii)リポソーム仲介形質移入、(iv)受容体仲介遺伝子導入、及び(v)DNAトランスポソンの使用(Morgan, R. A. 及び W. F. Anderson (1993) Annu. Rev. Biochem. 62:191-217、Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. 及び H. Recipon (1998) Curr. Opin. Biotechnol. 9:445-450)がある。

【0165】

ATRSの発現に影響を及ぼし得る発現ベクターには、限定するものではないが、PCDNA3.1、EPITAG、PRCCMV2、PREP、PVAXベクター(Invitrogen, Carlsbad CA)、PCMV-SCRIPT、PCMV-TAG、PEGSH/PERV(Stratagene, La Jolla CA)、PTET-OFF、PTET-ON、PTRE2、PTRE2-LUC、PTK-HYG(Clontech, Palo Alto CA)が含まれる。ATRSを発現させるために、(i)恒常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 - アクチン遺伝子等)、(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. 及び 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. 及び 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及び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/ミフェプリストーン誘導性プロモーター(Rossi, F. M. V. 及び H. M. Blau, 前出)、または(iii)正常な個体に由来するATRSをコードする内在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。

【0166】

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、当業者は経験にそれほど頼らないでもポリヌクレオチドを培養中の標的細胞に導入することが可能になる。別法では、リン酸カルシウム法(Graham, F. L. and A. J. Eb (1973) Virology 52:456-467)若しくは電気穿孔法(Neumann, B. ら (1982) EMBO J. 1:841-845)を用いて形質転換を行う。初代細胞にDNAを導入するためには、標準化された哺乳動物の形質移入プロトコルの修飾が必要である。

【0167】

10

20

30

40

50

本発明の別の実施例では、A T R Sの発現に関連する遺伝子欠損によって起こる疾患や異常症は、(i)レトロウイルス末端反復配列(L T R)プロモーター若しくは独立したプロモーターのコントロール下でA T R Sをコードするポリヌクレオチドと、(i i)好適なRNAパッケージングシグナルと、(i i i)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ディング配列を伴うRev応答性エレメント(R R E)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P F B及びP F B N E O)はS t r a t a g e n e社から市販されており、刊行データ(R i v i e r e , I . ら (1 9 9 5) P r o c . N a t l . A c a d . S c i . U . S . A . 9 2 : 6 7 3 3 - 6 7 3 7)に基づいている。上記データを引用することをもって本明細書の一部とする。ベクターは、好適なベクター産生細胞系(V P C L)において増殖され、V P C Lは、標的細胞上の受容体に対する向性を有するエンベロープ遺伝子またはV S V g等の乱交雑エンベロープタンパク質を発現する(A r m e n t a n o , D . ら (1 9 8 7) J . V i r o l . 6 1 : 1 6 4 7 - 1 6 5 0、B e n d e r , M . A . ら (1 9 8 7) J . V i r o l . 6 1 : 1 6 3 9 - 1 6 4 6、A d a m , M . A . 及び A . D . M i l l e r (1 9 8 8) J . V i r o l . 6 2 : 3 8 0 2 - 3 8 0 6、D u l l , T . ら (1 9 9 8) J . V i r o l . 7 2 : 8 4 6 3 - 8 4 7 1、Z u f f e r e y , R . ら (1 9 9 8) J . V i r o l . 7 2 : 9 8 7 3 - 9 8 8 0)。R i g gに付与された米国特許第5,910,434号(" M e t h o d f o r o b t a i n i n g r e t r o v i r u s p a c k a g i n g c e l l l i n e s p r o d u c i n g h i g h t r a n s d u c i n g e f f i c i e n c y r e t r o v i r a l s u p e r n a t a n t ")は、レトロウイルスパッケージング細胞系を得るための方法について開示しており、これを引用することをもって本明細書の一部とする。レトロウイルスベクターの増殖、細胞集団(例えばC D 4 ⁺ T細胞)の形質導入、及び形質導入した細胞の患者への戻しは、遺伝子治療の分野では当業者に公知の方法であり、多数の文献に記載されている(R a n g a , U . ら (1 9 9 7) J . V i r o l . 7 1 : 7 0 2 0 - 7 0 2 9、B a u e r , G . ら (1 9 9 7) B l o o d 8 9 : 2 2 5 9 - 2 2 6 7、B o n y h a d i , M . L . (1 9 9 7) J . V i r o l . 7 1 : 4 7 0 7 - 4 7 1 6、R a n g a , U . ら (1 9 9 8) P r o c . N a t l . A c a d . S c i . U . S . A . 9 5 : 1 2 0 1 - 1 2 0 6、S u , L . (1 9 9 7) B l o o d 8 9 : 2 2 8 3 - 2 2 9 0)。

【0168】

別法では、アデノウイルス系遺伝子治療の送達系を用いて、A T R Sの発現に関連する1つ或いは複数の遺伝子異常を有する細胞にA T R S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の作製及びパッケージングについては、当業者に公知である。複製欠損型アデノウイルスベクターは、免疫調節タンパク質をコードする遺伝子を脾臓の無損傷の脾島内に導入するために可変性であることが証明された(C s e t e , M . E . ら (1 9 9 5) T r a n s p l a n t a t i o n 2 7 : 2 6 3 - 2 6 8)。使用できる可能性のあるアデノウイルスベクターは、A r m e n t a n oに付与された米国特許第5,707,618号(" A d e n o v i r u s v e c t o r s f o r g e n e t h e r a p y ")に記載されており、引用することをもって本明細書の一部とする。アデノウイルスベクターについては、A n t i n o z z i , P . A . ら (1 9 9 9) A n n u . R e v . N u t r . 1 9 : 5 1 1 - 5 4 4、V e r m a , I . M . 及び N . S o m i a (1 9 9 7) N a t u r e 1 8 : 3 8 9 : 2 3 9 - 2 4 2も参照されたい。両文献は、引用することをもって本明細書の一部とする。

【0169】

別法では、ヘルペス系遺伝子治療の送達系を用いて、A T R Sの発現に関連する1つ或いは複数の遺伝子異常を有する標的細胞にA T R Sをコードするポリヌクレオチドを送達する。単純疱疹ウイルス(H S V)系のベクターは、H S V親和性の中枢神経細胞にA T R Sを導入する際に特に重要である。ヘルペス系ベクターの作製及びパッケージングは、当

業者に公知である。複製適格性単純ヘルペスウイルス(HSV)I型系のベクターは、レポーター遺伝子を霊長類の眼に送達するために用いられてきた(Liu, X. ら (1999) Exp. Eye Res. 169:385395を参照)。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第5,804,413号("Herpes simplex virus swains for gene transfer")に開示されており、該特許の引用をもって本明細書の一部とする。米国特許第5,804,413号には、ヒト遺伝子治療を含む目的のために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つの外在性遺伝子を有するゲノムを含む組換えHSV d92についての記載がある。上記特許はまた、ICP4、ICP27及びICP22のために除去される組換えHSV系統の作製及び使用について開示している。HSVベクターについては、Goins, W.F. ら (1999) J. Virol. 73:519-532、Xu, H. ら (1994) Dev. Biol. 163:152-161も参照されたい。両文献は、引用をもって本明細書の一部とする。クローン化ヘルペスウイルス配列の操作、巨大ヘルペスウイルスのゲノムの異なった部分を含む多数のプラスミドを形質移入した後の組換えウイルスの継代、ヘルペスウイルスの成長及び増殖、並びにヘルペスウイルスの細胞への感染は、当業者に公知の技術である。

10

20

30

40

50

【0170】

別法では、ウイルス(正の一本鎖RNAウイルス)ベクターを用いてATRSをコードするポリヌクレオチドを標的細胞に送達する。プロトタイプのウイルスであるセムリキ森林熱ウイルス(Semliki Forest Virus, 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がSFVゲノムに基づいていることが分かった(Garoff, H. 及び K.-J. Li (1998) Cun. Opin. Biotech. 9:464-469)。ウイルスRNAの複製中に、通常はウイルスカプシドタンパク質をコードするサブゲノムRNAが作り出される。このサブゲノムRNAは、完全長のゲノムRNAより高いレベルに複製されるため、酵素活性(例えばプロテアーゼ及びポリメラーゼ)を有するウイルスタンパク質に比べてキャプシドタンパク質が過剰産生される。同様に、ATRSをコードする配列を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ATRSをコードするRNAが産生され、高いレベルでATRSが合成される。通常はウイルスの感染が数日以内での細胞溶解に関係する一方で、シンドビスウイルス(SIN)の変異体を有するハムスター正常腎臓細胞(BHK-21)の持続的な感染を確立する能力は、ウイルスの溶解複製を遺伝子治療に適用できるように好適に変更可能であることを示唆している(Dryga, S.A. ら (1997) Virology 228:74-83を参照)。様々な宿主にウイルスを導入できることから、様々なタイプの細胞にATRSを導入することできる。或る集団におけるサブセットの細胞の特定形質導入は、形質導入前に細胞の選別を必要とし得る。ウイルスの感染性cDNAクローンの処置方法、ウイルスのcDNA及びRNAの形質移入方法及びウイルスの感染方法は、当業者に公知である。

【0171】

転写開始部位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。転写開始部位とは例えば開始部位から数えて約-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ような二重らせんの能力を阻害するので、三重らせん塩基対形成は有用である。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(Gee, J.E. ら (1994) in: Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY, 163-177ページ等を参照)。相補配列またはアンチセンス分子もまた、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

の翻訳を阻止するべく設計することができる。

【0172】

リボザイムは酵素性RNA分子であり、RNAの特異的切断を触媒するためにリボザイムを用いることもできる。リボザイム作用のメカニズムは、ヌクレオチド鎖切断に先立つ相補的標的RNAへのリボザイム分子の配列特異性ハイブリダイゼーションに参与している。例えば、ATRSをコードする配列のヌクレオチド鎖切断を、特異的且つ効果的に触媒する組換え型のハンマーヘッド型リボザイム分子が含まれる。

【0173】

任意の潜在的RNA標的内の特異的リボザイム切断部位は、GUA、GUU、GUC配列を含めたリボザイム切断部位に対する標的分子をスキャンすることによって先ず同定される。一度同定されると、オリゴヌクレオチドを機能不全にするような2次構造の特徴に対して切断部位を含む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とのハイブリダイゼーションの実施容易性をテストすることによって行うことができる。

10

【0174】

本発明の相補リボ核酸分子及びリボザイムは、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。任意の方法には、固相ホスホラミダイト化合物等のオリゴヌクレオチドを化学的に合成する方法がある。或いは、ATRSをコードするDNA配列の*in vitro*及び*in vivo*転写によってRNA分子を産出し得る。このようなDNA配列は、T7やSP6等の好適な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組み入れることが可能である。或いは、相補的RNAを構成的或いは誘導的に合成するようなこれらcDNA産物を、細胞系、細胞または組織内に導入することができる。

20

【0175】

細胞内の安定性を高め、半減期を長くするためにRNA分子を修飾することができる。限定するものではないが可能な修飾には、分子の5'末端、3'末端、あるいはその両方においてフランキング配列を追加したり、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ネートまたは2' Oメチルを使用したりすることが含まれる。この概念は、PNAの産出に固有のものであり、これら全ての分子に拡大することができる。それには、内在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、及びウリジンにアセチル-、メチル-、チオ-及び同様の修飾をしたものに加えて、非従来型塩基、例えばイノシン、クエオシン(queosine)、ワイプトシン(wybutosine)等を包含することによる。

30

【0176】

本発明の更なる実施例は、ATRSをコードするポリヌクレオチドの発現の変化に有効な化合物をスクリーニングする方法を含む。限定するものではないが特異ポリヌクレオチドの変異発現に有効な化合物には、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、転写因子その他のポリペプチド転写制御因子、及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変異し得る。従って、ATRS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、ATRS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、ATRS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、ATRS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。

40

【0177】

特異ポリヌクレオチドの変異発現における有効性に対して、少なくとも1個から複数個の試験化合物をスクリーニングし得る。試験化合物は、当分野で通常知られている任意の方法により得られる。このような方法には、ポリヌクレオチドの発現を変異させる場合と

50

、既存の、市販のまたは専売の、天然または非天然の化合物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及び/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に有効であることが知られているような化合物の化学修飾がある。A T R Sをコードするポリヌクレオチドを含むサンプルは、少なくとも1つの試験化合物に曝露して得る。サンプルには例えば、無傷細胞、透過化処理した細胞、無細胞再構成系または再構成生化学系があり得る。A T R Sをコードするポリヌクレオチドの発現における変化は、当分野で通常知られている任意の方法でアッセイする。通常、A T R S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有するプローブを用いたハイブリダイゼーションにより、特異ヌクレオチドの発現を検出する。ハイブリダイゼーション量を定量し、それによって1若しくは複数の試験化合物に曝露される及び曝露されないポリヌクレオチドの発現の比較に対する基礎を形成し得る。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変異する際に試験化合物が有効であることを示している。特異ポリヌクレオチドの変異発現に有効な化合物に対して、例えば Schizosaccharomyces pombe 遺伝子発現系 (Atkins, D. ら (1999) 米国特許第5,932,435号、Arndt, G.M. ら (2000) Nucleic Acids Res. 28:E15) または HeLa 細胞等のヒト細胞系 (Clarke, M.L. ら (2000) Biochem. Biophys. Res. Commun. 268:8-13) を用いてスクリーニングを実行する。本発明の特定の実施例は、特異的ポリヌクレオチド配列に対するアンチセンス活性のための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、修飾オリゴヌクレオチド) の組み合わせライブラリをスクリーニングすることに関与している (Bruice, T.W. ら (1997) 米国特許第5,686,242号、Bruice, T.W. ら (2000) 米国特許第6,022,691号)。

10

20

30

【0178】

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 及び 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一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、リボソーム注入または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 (Goldman, C.K. ら (1997) Nat. Biotechnol. 15:462-466 等を参照)。

【0179】

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、サル等の哺乳動物を含めて治療が必要な全ての対象に適用できる。

【0180】

本発明の追加実施例は、通常薬剤として許容できる賦形剤で処方される活性成分を有する成分の投与に関連する。賦形剤には例えば、セルロース、ゴム及びタンパク質がある。様々な処方が通常知られており、詳細は Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA) の最新版に記載されている。このような組成物は、A T R S、A T R S の抗体、擬態、アゴニスト、アンタゴニスト、または A T R S のインヒビターなどからなる。

40

【0181】

本発明に用いられる成分は、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。

【0182】

肺から投与する成分は、液状または乾燥粉末状で調製し得る。このような医薬品成分は通常、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重量有機薬) の場合には、速効製剤のエアロゾル輸送は当分野で公知である。高分子 (例えばより

50

大きなペプチド及びタンパク質)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輸送が最近向上したことにより、インスリン等の薬剤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. ら, 米国特許第5,997,848号等を参照)。肺送達は、針注射なしに投与する点で優れており、潜在的に有毒な浸透エンハンサーの必要性をなくす。

【0183】

本発明での使用に適した成分には、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成分が含まれる。有効投与量の決定は、当業者の能力の範囲内で行う。

【0184】

A T R Sまたはその断片を含む高分子を直接細胞内に輸送するべく、特殊な形態に組成物が調製されるのが好ましい。例えば、細胞不透過性高分子を含むリポソーム製剤は、細胞融合及び高分子の細胞内輸送を促進し得る。別法では、A T R Sまたはその断片をH I V T a t - 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は、マウスモデル系の脳を含む全ての組織の細胞に形質導入することがわかっている(Schwarze, S. R. ら (1999) Science 285: 1569 - 1572 参照)。

10

【0185】

任意の化合物に対して、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、或いは、動物モデル、例えばマウス、ウサギ、イヌまたはブタ等において、先ず治療の有効投与量を推定することができる。動物モデルはまた、好適な濃度範囲及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及び投与経路を決定することができる。

20

【0186】

医学的に効果的な薬用量は、症状や容態を回復させる、たとえばA T R Sまたはその断片、A T R Sの抗体、A T R Sのアゴニストまたはアンタゴニスト、インヒビターなどの活性処方成分の量に関連する。薬用有効度及び毒性は、細胞培養または動物実験における標準的な薬剤手法によって、例えばE D₅₀ (集団の50%の医薬的有效量)またはL D₅₀ (集団の50%の致死量)を測定するなどして決定することができる。毒性効果の治療効果に対する投与量の比は、治療指数であり、L D₅₀ / E D₅₀ 比として表すことができる。高い治療指数を示すような成分が望ましい。細胞培養アッセイ及び動物実験から得られたデータは、ヒトに用いるための投与量の範囲を調剤するのに用いられる。このような組成物が含まれる投与量は、毒性を殆ど或いは全く含まず、E D₅₀ 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、患者の感受性及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。

30

【0187】

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。効果的なレベルの活性成分を与え、或いは所望の効果を維持するべく、投与量及び投与を調節する。被験者に関する要素としては、疾患の重症度、患者の通常健康状態、患者の年齢、体重及び性別、投与の時間及び頻度、薬剤の配合、反応感受性及び治療に対する応答等を考慮する。作用期間が長い成分は、特定の製剤の半減期及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、或いは2週間に1度の間隔で投与し得る。

40

【0188】

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。特定の投与量及び輸送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、タンパク質またはインヒビターに対する処方とは異なる、ヌクレオチドに対す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、状態、位置等に特異的なものとなる。

【0189】

50

(診断)

別の実施例では、A T R S に特異的に結合する抗体が、A T R S の発現によって特徴付けられる疾患の診断、またはA T R S やA T R S 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られる。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。A T R S の診断アッセイには、抗体及び標識を用いてヒトの体液或いは細胞や組織から採取されたものからA T R S を検出する方法が含まれる。抗体は、修飾して或いは修飾しないで使用し、レポーター分子の共有結合性或いは非共有結合性の接着によって標識化し得る。多様なレポーター分子が本技術分野で知られており、それらを用いることができる。幾つかのレポーター分子については上記した。

10

【 0 1 9 0 】

A T R S を測定するためのE L I S A , R I A , 及びF A C Sを含む種々のプロトコルは、当分野では周知であり、変わった或いは異常なレベルのA T R S の発現を診断する元となるものを提供する。正常或いは標準的なA T R S の発現の値は、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者から採取した体液または細胞とA T R S に対する抗体とを結合させることによって決定する。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。被験者のA T R S の発現の量、制御及び疾患、生検組織からのサンプルが基準値と比較される。基準値と対象との偏差が疾患を診断するパラメータとなる。

【 0 1 9 1 】

本発明の別の実施例によれば、A T R S をコードするポリヌクレオチドを診断のために用いることもできる。用いられることができ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R N A 及びD N A 分子、そしてP N A が含まれる。このポリヌクレオチドを用いて、疾患と相関し得るA T R S を発現する生検組織における遺伝子の発現を検出し定量する。この診断アッセイを用いて、A T R S の存在の有無、更に過剰な発現を調べ、治療中のA T R S 値の調節を監視する。

20

【 0 1 9 2 】

一実施形態では、A T R S または近縁の分子をコードする遺伝子配列を含むポリヌクレオチド配列を検出可能なP C R プローブを用いたハイブリダイゼーションによって、A T R S をコードする核酸配列を同定することが可能である。例えば5'調節領域である高度に特異的な領域か、例えば保存されたモチーフであるやや特異性の低い領域から作られているかのプローブの特異性と、ハイブリダイゼーション或いは増幅のストリンジェントは、プローブがA T R S をコードする自然界の配列のみを同定するかどうか、或いは対立遺伝子や関連配列コードする自然界の配列のみを同定するかどうかによって決まるであろう。

30

【 0 1 9 3 】

プローブはまた、関連する配列の検出に利用され、A T R S をコードする任意の配列と少なくとも50%の配列同一性を有し得る。目的の本発明のハイブリダイゼーションプローブには、D N A あるいはR N A が可能であり、S E Q I D N O : 5 - 8 の配列、或いはA T R S 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。

40

【 0 1 9 4 】

A T R S をコードするD N A に対して特異的なハイブリダイゼーションプローブの作製方法には、A T R S 及びA T R S 誘導体をコードするポリヌクレオチド配列をm R N A プローブの作製のためのベクターにクローニングする方法がある。m R N A プローブ作製のためのベクターは、当業者に知られており、市販されており、好適なR N A ポリメラーゼ及び好適な標識されたヌクレオチドを加えることによって、i n v i t r oでR N A プローブを合成するために用いられ得る。ハイブリダイゼーションプローブは、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等の放射性核種、或いはアビジン/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等の酵素標識などが挙げられる。

50

【0195】

A T R Sをコードするポリヌクレオチド配列を用いて、A T R Sの発現に関連する疾患を診断することが可能である。限定するものではないがこのような疾患の例として細胞増殖異常も含まれ、その中には日光性角化症、動脈硬化症、アテローム性動脈硬化症、滑液包炎、硬変、肝炎、混合型結合組織病（M C T D）、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、原発性血小板血症と、腺癌、白血病、リンパ腫、黒色腫、骨髓腫、肉腫、奇形癌を含む癌、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、子宮の癌等が含まれる。また自己免疫/炎症疾患も含まれ、その中には後天性免疫不全症候群（A I D S）、アジソン病、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症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫多発性内分泌腺障害症（A P E C E D）、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、リンパ球毒素性一時的リンパ球減少症、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、過敏性大腸症候群、多発性硬化症、重症筋無力症、心筋または心膜の炎症、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、全身性紅斑性狼瘡、全身性硬化症、血小板減少症、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌の合併症、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、蠕虫の感染症及び外傷が含まれる。M A Dをコードするポリヌクレオチド配列は、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、P C R法、ディップスティック（d i p s t i c k）、ピン（p i n）、E L I S A式アッセイ、及び変異A T R Sの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マイクロアレイに使用することが可能である。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

【0196】

ある実施態様では、A T R Sをコードするヌクレオチド配列は、関連する疾患、特に上記した疾患を検出するアッセイにおいて有用であろう。A T R S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或いは組織のサンプルに加えることができるであろう。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、対照サンプルと較べて著しく変わっている場合は、サンプル内のA T R Sをコードするヌクレオチド配列の変異レベルにより、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推定するため、或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。

30

【0197】

A T R Sの発現に関連する疾患の診断の基準となるものを提供するために、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或いは増幅に好適な条件下で、動物或いはヒトの何れかの正常な被験者から抽出された体液或いは細胞と、A T R Sをコードする配列或いはその断片とを結合させ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用いて行った実験から得た値を正常な対象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。

40

【0198】

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通常ベースで繰り返し得る。連続アッセイから得られた結果を用いて、数日か

50

ら数ヶ月の期間にわたる治療の効果を示し得る。

【0199】

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、実際に臨床的症狀が現れる前に疾患を検出する方法を提供したりし得る。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。

【0200】

A T R Sをコードする配列から設計されたオリゴヌクレオチドのさらなる診断への利用には、P C Rの利用が含まれ得る。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、或いは*i n v i t r o*で産出し得る。オリゴマーは、好ましくはA T R Sをコードするポリヌクレオチドの断片、或いはA T R S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ト条件下で、近縁のD N A或いはR N A配列の検出、定量、或いはその両方のため用いることが可能である。

10

【0201】

或る実施態様において、A T R Sをコードするポリヌクレオチド配列由来のオリゴヌクレオチドプライマーを用いて、一塩基多型（S N P）を検出し得る。S N P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、挿入及び欠失である。限定するものではないがS N Pの検出方法には、制限酵素切断法（S S C P）及び蛍光S S C P（f S S C P）がある。S S C Pでは、A T R Sをコードするポリヌクレオチド配列由来の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（P C R）でD N Aを増幅する。D N Aは例えば、病変組織または正常組織、生検サンプル、体液その他に由来し得る。D N A内のS N Pは、一本鎖形状のP C R生成物の2次及び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。f S S C P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D N Aシーケンシング機などの高処理機器でアンプリマー（a m p l i m e r）の検出が可能になる。更に、インシリコS N P（i n s i l i c o S N P , i s S N P）と呼ばれる配列データベース分析法は、一般的なコンセンサス配列に配列されるような個々の重畳するD N A断片の配列を比較することにより、多形性を同定し得る。これらのコンピュータベースの方法は、D N Aの実験室での調整及び統計モデル及びD N A配列クロマトグラムの自動分析を用いたシーケンシングのエラーに起因する配列の変異をフィルタリングして除去する。別の態様では、例えばハイスループットM A S S A R R A Yシステム（S e q u e n o m , I n c . , S a n D i e g o C A）を用いた質量分析によりS N Pを検出し、特徴付ける。

20

30

【0202】

A T R Sの発現を定量するために用い得る方法には、ヌクレオチドの放射標識またはバイオチン標識、調節核酸の相互増幅（c o a m p l i f i c a t i o n）及び標準的な曲線に結果が加えられたものが含まれる（M e l b y , P . C . 他（1993）J . I m m u n o l . M e t h o d s 159:235244; D u p l a a , C . 他（1993）A n a l . B i o c h e m . 212:229236を参照）。目的のオリゴマーが種々の希釈液中に存在し、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のアッセイを行うことによって、複数のサンプルの定量速度を加速することができる。

40

【0203】

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由来のオリゴヌクレオチドまたはより長い断片を、マイクロアレイにおける標的として用いることができる。多数の遺伝子の関連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

50

はまた、遺伝変異体、突然変異及び多形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剤の活性を開発及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、有効的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有効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。

【0204】

別の実施例では、A T R S、A T R Sの断片、A T R Sに特異的な抗体をマイクロアレイ上のエレメントとして用いることができる。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、薬剤 - 標的相互作用及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。

10

【0205】

或る実施例は、或る組織または細胞タイプの転写イメージを生成するような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り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及び相対存在量を定量することにより分析し得る (Seilliamer らの米国特許第5,840,484号 "Comparative Gene Transcript Analysis" を参照。当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)。従って、特定の組織または細胞タイプの転写または逆転写全体に本発明のポリヌクレオチドまたはその補体をハイブリダイズすることにより、転写イメージを生成し得る。或る実施例では、本発明のポリヌクレオチドまたはその補体がマイクロアレイ上のエレメントのサブセットを複数含む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。

20

【0206】

転写イメージは、組織、株化細胞、生検またはその生物学的サンプルから単離した転写物を用いて生成し得る。転写イメージは従って、組織または生検サンプルの場合には in vivo、または株化細胞の場合には in vitro での遺伝子発現を反映する。

【0207】

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、工業的または天然の環境化合物の毒性試験のみならず、in vitro モデル系及び薬剤の前臨床評価に関連して使用し得る。全ての化合物は、作用及び毒性のメカニズムを暗示する、しばしば分子フィンガープリントまたは毒性サインと称されるような特徴的な遺伝子発現パターンを惹起する (Nuwaysir, E. F. ら (1999) Mol. Carcinog. 24:153-159、Steiner, S. 及び N. L. Anderson (2000) Toxicol. Lett. 112-113:467-471、当該文献は特に引用することを以って本明細書の一部となす)。試験化合物が、既知の毒性を有する化合物のサインと同一のサインを有する場合には、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サインは、多数の遺伝子及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は、最も有用且つ正確である。理想的には、発現のゲノム全域にわたる測定が最高品質のサインを提供する。自己の発現が任意の試験された化合物により変化しない遺伝子が同様に重要であっても、このような遺伝子の発現レベルを用いて残りの発現データを規準化する。規準化手順は、異なる化合物で処理した後の発現データの比較に有用である。毒性サインの要素への遺伝子機能を割り当てることが毒性メカニズムの阻止に役立つが、毒性の予測につながるサインの統計的に一致させるのに遺伝子機能の知識は必要とされない (例えば2000年2月29日に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>

30

40

50

．htmlで入手可能である）。従って、中毒学的スクリーニングの際に毒性サインを用いて、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ことである。

【0208】

或る実施例では、核酸を含有する生物学的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。処理した生物学的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つ若しくは複数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。処理した生物学的サンプル中の転写レベルを、非処理生物学的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、処理されたサンプル中で試験化合物が引き起こす毒性反応を示す。

【0209】

別の実施例は、本発明のポリペプチド配列を用いて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連する。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、個々に更に分析の対象とすることができ、プロテオーム発現パターン即ちプロファイルは、所与の条件下で所与の時間に発現したタンパク質の数及び相対存在量を定量することにより分析し得る。従って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及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 and Anderson）。タンパク質は、通常クーマシーブルーまたはシルバーまたは蛍光染色などの物質を用いてゲルで染色することにより、分散した、独特な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非処理のいずれかの生物学的サンプルから得られる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ケンシングする。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が得られる。

【0210】

プロテオームのプロファイルは、ATRSに特異的な抗体を用いてATRS発現レベルを定量することによっても作成可能である。或る実施例では、マイクロアレイ上でエレメントとして抗体を用い、マイクロアレイをサンプルに曝して各アレイ要素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. ら（1999）Anal. Biochem. 270:103-111、Mendoza, L. G. ら（1999）Biotechniques 27:778-788を参照）。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、チオールまたはアミノ反応性蛍光化合物を用いてサンプル中のタンパク質を反応させ、各アレイのエレメントにおける蛍光結合の量を検出し得る。

【0211】

プロテオームレベルでの毒性サイン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サインと並行に分析するべきである。或る組織の或るタンパク質に対しては、転写とタンパク質の存在量の相関が乏しいこともあるので（Anderson, N. L. and J. Seilhamer（1997）Electrophoresis 18:533-537）、転写イメージにはそれ程影響しないがタンパク質のプロファイルを変化させるような化合物の分析においてプロテオーム毒性サインは有用たり得る。更に、体液中での転写の分析はmRNA急速な分解により困難であるので、タンパク質のプロファイル作成はこのような場合により信頼でき、情報価値がある。

【0212】

別の実施例では、タンパク質を含有する生物学的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。処理された生物学的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、非処理生物学的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量の差は、処理サンプル中の試験化合物に対する反応を示す。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。

【0213】

別の実施例では、タンパク質を含有する生物学的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。生物学的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。処理された生物学的サンプル中のタンパク質の量を、非処理生物学的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量の差は、処理サンプル中の試験化合物に対する反応を示す。

10

【0214】

マイクロアレイは、本技術分野でよく知られている方法を用いて調製し、使用し、そして分析する(Brennan, T. M. ら (1995) の米国特許第5,474,796号、Schena, M. ら (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619、Baldeschweiler らの (1995) PCT出願第WO95/251116号、Shalon, D. らの (1995) PCT出願第WO95/35505号、Heller, R. A. ら (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155、Heller, M. J. らの (1997) 米国特許第5,605,662号等を参照)。様々なタイプのマイクロアレイが周知であり、詳細については、DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, Londonに記載されている。該文献は、特別に引用することを以て本明細書の一部となす。

20

【0215】

本発明の別の実施例ではまた、ATRSをコードする核酸配列を用いて、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、或る例では、コード配列全体で非コード配列が好ましい。例えば、多重遺伝子ファミリー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。核酸配列は、特定の染色体、染色体の特定領域または人工形成の染色体、例えば、ヒト人工染色体(HAC)、酵母人工染色体(YAC)、細菌人工染色体(BAC)、細菌P1産物、或いは単一染色体cDNAライブラリに対してマッピングされる(Harrington, J. J. 他 (1997) Nat. Genet. 15:345-355; Price, C. M. (1993) Blood Rev. 7:127134; Trask, B. J. (1991) Trends Genet. 7:149154を参照)一度マッピングすると、本発明の核酸配列を用いて例えば病状の遺伝を特定の染色体領域の遺伝または制限酵素断片長多型(RFLP)と関連させるような遺伝子連鎖地図を発生させ得る。(例として、Lander, E. S. 及び D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357を参照)

30

40

蛍光原位置ハイブリッド形成法(FISH)は、他の物理的及び遺伝地図データと関連し得る(前出Meyers, 965-968ページのHeinz-Ulrich, 他 (1995)を参照)。遺伝地図データの例は、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ることができる。物理的な染色体地図上のATRSをコードする遺伝子の位置と特定の疾患との相関性、或いは特定の疾患に対する素因が、このような疾患と関連するDNA領

50

域の決定に役立つため、更なる位置を決定するクローニングが行われる。

【0216】

確定した染色体マーカーを用いた結合分析等の物理的マッピング技術及び染色体標本原位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。例えばマウスなどの別の哺乳動物の染色体上に遺伝子を配置することにより、特定のヒト染色体の数或いはアームが分かっていない場合でも関連するマーカーを明らかにし得る。この情報は、位置クローニングその他の遺伝子発見技術を用いて遺伝的疾患を調査する研究者にとって価値がある。疾患または症候群が、血管拡張性失調症の11q22-23領域等、特定の遺伝子領域への遺伝的結合によって大まかに位置決めがなされると、該領域に対するいかなるマッピングも、更なる調査のための関連遺伝子或いは調節遺伝子を表すことができる（Ga 10
t t i , R . A . e t a l (1 9 8 8) N a t u r e 3 3 6 : 5 7 7 5 8 0 を参照）。転座、反転等に起因する、健常者、保有者、感染者の三者間における染色体位置の相違を発見するために本発明のヌクレオチド配列を用いてもよい。

【0217】

本発明の別の実施例では、A T R S、その触媒作用断片或いは免疫原断片またはそのオリゴペプチドを、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のスクリーニングに用いることができる。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になる。A T R Sと検査する薬剤との結合による複合体の形成を測定してもよい。

【0218】

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（G e y s e n , 他 (1 9 8 4) P C T 出願番号 W O 8 4 / 0 3 5 6 4 を参照）この方法においては、多数の異なる小さな試験用化合物を固体基質上で合成する。試験用化合物は、A T R S、或いはその断片と反応してから洗浄される。次に、本技術分野でよく知られている方法で、結合したA T R Sを検出する。精製したA T R Sはまた、上記した薬剤のスクリーニング技術において用いるプレート上で直接コーティングすることもできる。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。

【0219】

別の実施例では、A T R Sと結合可能な中和抗体がA T R Sと結合するため試験用化合物と特に競合する、競合的薬剤スクリーニングアッセイを用いることができる。この方法では、抗体が、A T R Sと1つ以上の抗原決定因子を共有するどのペプチドの存在も検出する。

【0220】

別の実施例では、発展途上の分子生物学技術にA T R Sをコードするヌクレオチド配列を用いて、限定はされないが、現在知られているトリプレット暗号及び特異的な塩基対相互作用などのヌクレオチド配列の特性に依存する新しい技術を提供することができる。

【0221】

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きるであろう。従って、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 40
にも本発明を限定するものではない。

【0222】

前述した及び以下に記載する全ての特許出願、特許、刊行物、特に米国特許出願第60 / 207, 248号、同第60 / 208, 791号および同第60 / 210, 585号に言及することをもって本明細書の一部とする。

【0223】

(実施例)

1 cDNAライブラリの作製

I n c y t e c D N A は、L I F E S E Q G O L D データベース（I n c y t e G e n o m i c s , P a l o A l t o C A ）に記載されたcDNAライブラリに由来する 50

ものであり、表4の列5に列記した。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た組織もあり、また、ホモジナイズしてフェノールまたは好適な変性剤の混合液に溶解した組織もある。変性剤の混合液は、例えばフェノールとグアニジニウムイソチオシアネートの単相溶液であるTRIzol (Life Technologies) 等である。結果として得られた溶解物は、塩化セシウムにおい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、或いは別の方法を用いて、溶解物からRNAを沈殿させた。

RNAの純度を高めるため、RNAのフェノールによる抽出及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, Valencia 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。

10

20

30

【0224】

場合によってはStratagene社へのRNA提供を行い、対応するcDNAライブラリをStratagene社が作製することもあった。そうでない場合は、本技術分野で公知の推奨方法または類似の方法を用いて、UNIZAPベクターシステム(Stratagene)またはSUPERSCRIPTプラスミドシステム(Life Technologies)を用いてcDNAを合成し、cDNAライブラリを作製した。(前出のAusubel, 1997, unit 5.1-6.6等を参照。)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、好適な制限酵素でcDNAを消化した。殆どのライブラリに対して、cDNAのサイズ(300~1000bp)選択は、SEPHACRYL S1000、SEPHAROSE CL2BまたはSEPHAROSE CL4Bカラムクロマトグラフィー(Amersham Pharmacia Biotech)、或いは調製用アガロースゲル電気泳動法を用いて行った。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でcDNAを消化した。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)またはpINCY(Incyte Pharmaceuticals, Palo Alto CA)等である。組換えプラスミドは、Stratagene社のXL1-Blue、XL1-Blue MRFまたはSOLR、或いはLife Technologies社のDH5、DH10BまたはELECTROMAX DH10Bを含む大腸菌細胞に形質転換した。

【0225】

2 cDNAクローンの単離

UNIZAPベクターシステム(Stratagene)を用いた*in vivo*切除によって、或いは細胞溶解によって、プラスミドを宿主細胞から回収した。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 8 Plasmid、QIAWELL 8 Plus Plasmid及びQIAWELL 8 Ultra Plasmid精製システム、R.E.A.L. Prep 96プラスミドキットの中から少なくとも1つを用いて、プラスミドを精製した。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。

40

【0226】

別法では、ハイスルーブットフォーマットにおいて直接結合PCR法を用いて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)。宿主細胞の溶解及び熱サイクリング過程は、単一

50

反応混合液中で行った。サンプルを処理し、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PICO GREEN色素 (Molecular Probes, Eugene OR) 及びFluoroskan II 蛍光スキャナー (Labsystems Oy, Helsinki, Finland) を用いて蛍光分析的に定量した。

【0227】

3 シークエンシング及び分析

実施例2に記載したようにプラスミドから回収したIncycyte 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、標準的方法或いは高処理装置、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 cycle sequencing ready reaction kit (Applied Biosystems) に与えられた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及び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プロトコル及び塩基対呼び出しソフトウェアを用いるABI PRISM 373または377シークエンシングシステム (Applied Biosystems) か、或いはその他の本技術分野でよく知られている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, unit 7.7に概説) を用いて決定した。cDNA配列の幾つかを選択して、実施例8に記載した方法で配列を伸長させた。

【0228】

cDNA配列に由来するポリヌクレオチド配列は、ベクター、リンカー及びポリ(A)配列の除去し、あいまいな塩基対をマスクすることによって有効性を確認した。その際、BLAST、動的プログラミング及び隣接ジヌクレオチド頻度分析に基づくアルゴリズム及びプログラムを用いた。次に、BLAST、FASTA及びBLIMPSに基づくプログラムを用いて、プログラム中の注釈を得るべく、公共のデータベース、例えばGenBankの霊長類及びげっ歯類、哺乳動物、脊椎動物、真核生物のデータベースと、BLOCKS、PRINTS、DOMO、PRODOM及びPFAM等のHidden Markov Model (HMM) ベースのタンパク質ファミリーデータベースの選択に対するIncycyte cDNA配列またはその翻訳を問い合わせた (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。Eddy, S. R. (1996) Cuff. Opin. Struct. Biol. 6:361-365等を参照)。問合せは、BLAST、FASTA、BLIMPS及びHMMERに基づくプログラムを用いて行った。Incycyte cDNA配列は、完全長のポリヌクレオチド配列を産出するように構築した。或いは、GenBank cDNA、GenBank EST、ステッチされた配列、ストレッチされた配列またはGenscan予測コード配列 (実施例4及び5を参照) を用いてIncycyte cDNAの集団を完全長まで伸長させた。Phred、Phrap及びConsedに基づくプログラムを用いて構築し、GenMark、BLAST及びFASTAに基づくプログラムを用いてcDNAの集団をオープンリーディングフレームに対してスクリーニングした。対応する完全長ポリペプチド配列を誘導するべく完全長ポリヌクレオチド配列を翻訳した。或いは、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。引き続いて、GenBankタンパク質データベース (genpept)、SwissProt、BLOCKS、PRINTS、DOMO、PRODOM及びPrositate等のデータベース、PFAM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に

対する問合せによって完全長ポリペプチド配列を分析した。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）及びLASERGENEソフトウェア（DNASTAR）を用いて分析した。ポリヌクレオチド及びポリペプチド配列アラインメントは、アラインメントした配列と配列の一致率も計算するMEGALIGNマルチシークエンスアラインメントプログラム（DNASTAR）に組み込まれているようなCLUSTALアルゴリズムによって特定されるデフォルトパラメータを用いて生成する。

【0229】

Incyte cDNA及び完全長配列の分析及びアセンブリに利用したツール、プログラム及びアルゴリズムの概略と、適用可能な説明、引用文献、閾値パラメータを表7に示す。用いたツール、プログラム及びアルゴリズムを表5の列1に、それらの簡単な説明を列2に示す。列3は好適な引用文献であり、全ての文献はそっくりそのまま引用を以って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、確率値その他のパラメータを示す（スコアが高ければ高いほど2配列間の相同性が高くなる）。

10

【0230】

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構築及び分析に用いる上記のプログラムは、SEQ ID NO: 5 - 8のポリヌクレオチド配列断片の同定にも利用できる。ハイブリダイゼーション及び増幅技術に有用である約20～約4000ヌクレオチドの断片を表4の列4に示した。

20

【0231】

4 ゲノムDNAからのコード配列の同定及び編集

推定上のアミノアシルtRNA合成酵素は、公共のゲノム配列データベース（例えば、gbpriやgbhtg）においてGenscan遺伝子同定プログラムを実行して初めに同定された。Genscanは、様々な生物からゲノムDNA配列を分析する汎用遺伝子同定プログラムである（Burge, C. and S. Karlin (1997) J. Mol. Biol. 268: 78 - 94 及びBurge, C. 及び S. Karlin (1998) Cuff. Opin. Struct. Biol. 8: 346 - 354 参照）。プログラムは予測エキソンを連結し、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、ポリヌクレオチド及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、30kbに設定した。これらのGenscan推定cDNA配列の内、どの配列がアミノアシルtRNA合成酵素をコードするかを決定するために、コードされたポリペプチドをPFAMモデルにおいてアミノアシルtRNA合成酵素について問合せて分析した。アミノアシルtRNA合成酵素としてアノテーションが付けられたIncyteのcDNA配列に対する相同性を基に、可能性のあるアミノアシルtRNA合成酵素が同定された。こうして選択されたGenscan予測配列は、次にBLAST分析により公共データベースgbpri及びgbhtgと比較した。必要であれば、genpeptからヒットしたトップのBLASTと比較することによりGenscan予測配列を編集し、余分なまたは取り除かれたエキソンなどのGenscanにより予測された配列のエラーを修正する。BLAST分析はまた、任意のIncyte cDNAまたはGenscan予測配列の公共cDNA適用範囲の発見に用いられるので、転写の証拠を提供する。Incyte cDNA適用範囲が利用できる場合には、この情報を用いてGenscan予測配列を修正または確認した。完全長ポリヌクレオチド配列は、実施例3に記載された構築プロセスを用いて、Incyte cDNA配列及び/または公共のcDNA配列でGenscan予測コード配列を構築することにより得た。或いは、完全長ポリヌクレオチド配列は編集または非編集のGenscan予測コード配列に完全に由来する。

30

40

【0232】

5 ゲノム配列データのcDNA配列データへのアセンブリ

50

ステッチ配列 (S t i c h e d S e q u e n c e)

部分 c D N A 配列は、実施例 4 に記載の G e n s c a n 遺伝子同定プログラムにより予測されたエキソンを用いて伸長させた。実施例 3 に記載されたように構築された部分 c D N A は、ゲノム D N A にマッピングし、関連する c D N A 及び 1 つ若しくは複数のゲノム配列から予測された G e n s c a n エキソンを含むクラスタに分解した。c D N A 及びゲノム情報を統合するべくグラフ理論及び動的プログラミングに基づくアルゴリズムを用いて各クラスタを分析し、引き続いて確認、編集または伸長して完全長配列を産出するような潜在的スプライス変異体を生成した。間隔全体の長さがクラスタ中の 2 以上の配列に存在するような配列を同定し、そのように同定された間隔は推移により等しいと考えられた。例えば、1 つの c D N A 及び 2 つのゲノム配列に間隔が存在する場合、3 つの間隔は全て等しいと考えられる。このプロセスは、無関係であるが連続したゲノム配列を c D N A 配列により結び合わせて架橋し得る。このようにして同定された区間を、親配列 (p a r e n t s e q u e n c e) に沿って現われるようにステッチアルゴリズムで縫い合わせ、可能な最も長い配列および変異配列を作製する。1 種類の親配列に沿って発生した間隔と間隔との連鎖 (c D N A - c D N A またはゲノム配列 - ゲノム配列) は、親の種類を変える連鎖 (c D N A - ゲノム配列) に優先した。結果として得られるステッチ配列は、B L A S T 分析により公共データベース g e n p e p t 及び g b p r i に翻訳されて比較された。G e n s c a n により予測された不正確なエキソンは、g e n p e p t からヒットしたトップの B L A S T と比較することにより修正した。必要な場合には、追加 c D N A 配列を用いるかゲノム D N A の検査により配列を更に伸長させた。

10

20

【 0 2 3 3 】

ストレッチ配列 (S t r e t c h e d S e q u e n c e)

部分 D N A 配列は、B L A S T 分析に基づくアルゴリズムにより完全長まで伸長された。まず、B L A S T プログラムを用いて、G e n B a n k の霊長類、げっ歯類、哺乳動物、脊椎動物及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アセンブルされた部分 c D N A を問い合わせた。次に、最も近い G e n B a n k タンパク質相同体を B L A S T 分析により I n c y t e c D N A 配列または実施例 4 に記載の G e n S c a n 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 (H S P) を用いてキメラタンパク質を産出し、翻訳した配列を G e n B a n k タンパク質相同体上にマッピングした。元の G e n B a n k タンパク質相同体に関連して、キメラタンパク質内で挿入または削除が起こり得る。G e n B a n k 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 D N A 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決定した。

30

【 0 2 3 4 】

6 A T R S をコードするポリヌクレオチドの染色体マッピング

S E Q I D N O : 5 - 8 を構築するために用いた配列を、B L A S T 及び S m i t h - W a t e r m a n アルゴリズムを用いて、I n c y t e L I F E S E Q データベース及び公共のドメインデータベースの配列と比較した。S E Q I D N O : 5 - 8 と一致するこれらのデータベースの配列を、P h r a p (表 7) などの構築アルゴリズムを使用して、連続及び重複した配列のクラスターに組み入れた。スタンフォード・ヒトゲノムセンター (S H G C) 、ホワイトヘッド・ゲノム研究所 (W I G R) 、G e n e t h o n 等の公的な情報源から入手可能な放射線ハイブリッド及び遺伝地図データを用いて、クラスタ化された配列が前もってマッピングされたかを測定した。マッピングされた配列がクラスタに含まれている結果、個々の配列番号を含めてそのクラスタの全配列が地図上の位置に割り当てられた。

40

【 0 2 3 5 】

地図上の位置は、ヒト染色体の範囲または間隔として表される。センチモルガン間隔の地図上の位置は、染色体の p アームの末端に関連して測定する。(センチモルガン (c M)

50

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 c Mは、ヒト中のDNAの1メガベース (M b) にほぼ等しい。尤も、この値は、組換えのホットスポット及びコールドスポットに起因して広範囲に変化する。) c M距離は、配列が各クラスタ内に含まれるような放射線ハイブリッドマーカーに対して境界を提供するようなGenethonによってマッピングされた遺伝マーカーに基づく。NCBI「GeneMap99」(<http://www.ncbi.nlm.nih.gov/genemap>)などの一般が入手可能なヒト遺伝子マップおよびその他の情報源を用いて、上記した区間が既に同定されている疾患遺伝子マップ内若しくは近傍に位置するかを決定できる。このようにして、SEQ ID NO: 5は97.1から116.6センチモルガン内で染色体12にマップされた。

10

【0236】

7 ポリヌクレオチド発現の分析

ノーザン分析は、転写された遺伝情報の存在を検出するために用いられる実験技術であり、標識されたヌクレオチド配列の、特定の細胞種または組織からのRNAが結合される膜へのハイブリダイゼーションに関与している。(前出のSambrook, 7章、同Ausubel, F.M. ら, 4章及び16章等を参照。)

【0237】

BLASTに適用する類似のコンピュータ技術を用いて、GenBankやLifeSeq (Incyte Pharmaceuticals) 等のヌクレオチドデータベースにおいて同一または関連分子を検索する。ノーザン分析は、多数の膜系ハイブリダイゼーションよりも断然速い。更に、特定の同一を厳密な或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変更することができる。検索の基準は積スコアであり、次式で定義される。

20

【0238】

【数1】

(BLASTスコア×配列一致率)

$5 \times (\text{長さ(配列1)}, \text{長さ(配列2)}) \text{の最小値}$

積スコアは、2つの配列間の類似度及び配列が一致する長さの両方を考慮している。積スコアは、0～100の規準化された値であり、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。高スコアリングセグメント対(HSP)に一致する各塩基に+5のスコアを割り当て、各不適性塩基対に-4を割り当てることにより、BLASTスコアを計算する。2つの配列は、2以上のHSPを共有し得る(ギャップにより離隔され得る)。2以上のHSPがある場合には、最高BLASTスコアの塩基対を用いて積スコアを計算する。積スコアは、断片的重畳と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%重畳しているか、他端が88%一致し、100%重畳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%重畳しているか、他端が79%一致し、100%重畳しているかのいずれかの場合に得られる。

30

40

【0239】

或いは、ATRSをコードするポリヌクレオチド配列は、由来する組織に対して分析する。例えば或る完全長配列は、Incyte cDNA配列(実施例3を参照)と少なくとも一部は重畳するようにアセンブルされる。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、以下の器官/組織カテゴリー即ち心血管系、結合組織、消化器系、胚構造、内分泌系、外分泌腺、女性生殖器、男性生殖器、生殖細胞、血液及び免疫系、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/混合性または尿路の1つに分類される。各カテゴリーのライブラリ数を数

50

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/病状カテゴリー即ち癌、細胞系、発達、炎症、神経性、外傷、心血管、鬱血、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。得られるパーセンテージは、A T R Sをコードするc D N Aの疾患特異的な発現を反映する。c D N A配列およびc D N Aライブラリ/組織の情報は、L I F E S E Q G O L D データベース (I n c y t e G e n o m i c s , P a l o A l t o C A) から得ることができる。

【0240】

8 A T R S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もまた、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を用いて該断片を伸長させて生成した。一方の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、他方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。開始プライマーの設計は、長さが約22~30ヌクレオチド、GC含有率が50%以上となり、約68~72の温度で標的配列にアニーリングするように、O L I G O 4 . 0 6ソフトウェア (N a t i o n a l B i o s c i e n c e s) 或いは別の適切なプログラムを用いて、c D N Aから設計した。ヘアピン構造及びプライマー-プライマー二量体を生ずるようなヌクレオチドの区間は全て回避した。

【0241】

配列を伸長するために、選択されたヒトc D N Aライブラリを用いた。2段階以上の伸長が必要または望ましい場合には、付加的プライマー或いはプライマーのネステッドセットを設計した。

【0242】

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P C R法によって得られた。P C Rは、P T C - 2 0 0 サーマルサイクラー (M J R e s e a r c h , I n c .) を用いて96穴プレート内で行った。反応混合液は、鋳型DNA及び200 nmolの各プライマー、 Mg^{2+} と $(NH_4)_2SO_4$ と-メルカプトエタノールを含むバッファー、Taq DNAポリメラーゼ (A m e r s h a m P h a r m a c i a B i o t e c h)、E L O N G A S E酵素 (L i f e T e c h n o l o g i e s)、Pfu DNAポリメラーゼ (S t r a t a g e n e) を含む。プライマーの組、P C I AとP C I Bに対して以下のパラメータで増幅を行った。ステップ 1: 94 で3分間、ステップ 2: 94 で15分間、ステップ 3: 60 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、及び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。別法では、プライマー対、T7とSK+に対して以下のパラメータで増幅を行った。ステップ 1: 94 で3分間、ステップ 2: 94 で15分間、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、及び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。

【0243】

各ウェルのDNA濃度は、1X T E及び0.5 μ lの希釈していないP C R産物に溶解した100 μ lのP I C O G R E E N定量試薬 (0.25 (v/v) P I C O G R E E N; M o l e c u l a r P r o b e s , E u g e n e O R) を不透明な蛍光光度計プレート (C o m i n g C o s t a r , A c t o n M A) の各ウェルに分配してDNAが試薬と結合できるようにして測定する。サンプルの蛍光を計測してDNAの濃度を定量するべくプレートをF l u o r o s k a n I I (L a b s y s t e m s O y , H e l s i n k i , F i n l a n d) でスキャンした。反応混合物のアリコート5~10 μ lを1%アガロースミニ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決定した。

【0244】

伸長させたヌクレオチドは、脱塩及び濃縮して384穴プレートに移し、C v i J I コレラウイルスエンドヌクレアーゼ (M o l e c u l a r B i o l o g y R e s e a r c h

10

20

30

40

50

、M a d i s o n W I) を用いて消化し、p U C 18ベクター (A m e r s h a m P h a r m a c i a B i o t e c h) への再連結反応前に音波処理またはせん断した。シ
ョットガン・シーケンシングのために、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A g a r A C E (P r o m
e g a) で消化した。伸長させたクローンをT 4リガーゼ (N e w E n g l a n d B i
o l a b s , B e v e r l y M A) を用いてp U C 18ベクター (A m e r s h a m
P h a r m a c i a B i o t e c h) に再連結し、P f u D N Aポリメラーゼ (S t r
a t a g e n e) で処理して制限部位の張出部 (o v e r h a n g) を満たし、コンピテ
ント大腸菌細胞に形質移入した。形質移入した細胞を選択して抗生物質を含む培地に移し
、それぞれのコロニーを切りとってL B / 2 Xカルベニシリン培養液の384ウェルプレ
ートに37 で一晩培養した。 10

【0245】

細胞を溶解して、T a q D N Aポリメラーゼ (A m e r s h a m P h a r m a c i a
B i o t e c h) 及びP f u D N Aポリメラーゼ (S t r a t a g e n e) を用いて以
下の手順で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分間、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、及び4を20回繰り返す。ステップ 6 : 72 で5分間、ステッ
プ 7 : 4 で保存。上記したようにP I C O G R E E N試薬 (M o l e c u l a r P r
o b e s) でDNAを定量化した。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、D Y E N A M I C エネルギートランスファー シークエンシングプライマー、及び
D Y E N A M I C D I R E C T k i t (A m e r s h a m P h a r m a c i a B i o
t e c h) またはA B I P R I S M B I G D Y E ターミネーターサイクル シークエン
シング反応キット (T e r m i n a t o r c y c l e s e q u e n c i n g r e a d
y r e a c t i o n k i t) (A p p l i e d B i o s y s t e m s) を用いてシー
クエンシングした。 20

【0246】

同様に、上記手順を用いて完全長ヌクレオチド配列を検証し、或いはそのような伸長のた
めに設計されたオリゴヌクレオチド及び適切なゲノムライブラリを用いて5'調節配列を
得る。 30

【0247】

9 個々のハイブリダイゼーションプローブの標識及び使用

S E Q I D N O : 5 - 8 から導き出されたハイブリダイゼーションプローブを用いて、
c D N A、m R N A、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
しても事実上同一の手順が用いられる。オリゴヌクレオチドは、O L I G O 4 . 0 6 ソフ
トウェア (N a t i o n a l B i o s c i e n c e s) 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、[γ - 32 P] アデノシン3リン酸 (A m e r
s h a m P h a r m a c i a B i o t e c h) 250 μ C i と、T 4 ポリヌクレオチド
キナーゼ (D u P o n t N E N , B o s t o n M A) を混合することにより標識する 40
。標識したオリゴヌクレオチドは、S E P H A D E X G - 25 超細繊維分子サイズ排除デ
キストランビードカラム (A m e r s h a m P h a r m a c i a B i o t e c h) を用
いて実質的に精製する。A s e I、B g l I I、E c o R I、P s t I、X b a I ま
たはP v u I I (D u P o n t N E N) 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
10⁷ 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。

【0248】

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 (N y t r
a n P l u s , S c h l e i c h e r & S c h u e l l , D u r h a m N H) に移
す。ハイブリダイゼーションは、40 で16時間行う。非特異的シグナルを除去する 50

ため、例えば0.1×クエン酸ナトリウム食塩水及び0.5%ドデシル硫酸ナトリウムに一致する条件下で、プロット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。

【0249】

10 マイクロアレイ

マイクロアレイの表面上でアレイエレメントの連鎖または合成は、フォトリソグラフィ、圧電印刷（インクジェット印刷、前出のBaldeschweiler等を参照）、機械的マイクロスポッティング技術及びこれらから派生したものをを用いて達成することが可能である。上記各技術において基質は、均一且つ非多孔性の固体とするべきである（Schena（1999）前出）。推奨する基質には、シリコン、シリカ、スライドガラス、ガラスチップ及びシリコンウエハがある。或いは、ドットプロット法またはスロットプロット法に類似のアレイを利用して、熱的、紫外線の、化学的または機械的結合手順を用いて基質の表面にエレメントを配置及び結合させてもよい。通常のアレイは、手作業で、または利用可能な方法や機械を用いて作製でき、任意の適正数のエレメントを有し得る（Schena, M. ら（1995）Science 270:467-470、Shallon, D. ら（1996）Genome Res. 6:639-645、Marshall, A. and J. Hodgson（1998）Nat. Biotechnol. 16:27-31.を参照）。

10

【0250】

完全長cDNA、発現遺伝子配列断片（EST）、またはその断片またはオリゴマーは、マイクロアレイのエレメントを構成し得る。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、レーザGENEソフトウェア（DNASTAR）等の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。アレイエレメントは、生物学的サンプル中でポリヌクレオチドを用いてハイブリダイズされる。生物学的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またはその他の分子タグに接合される。ハイブリダイゼーション後、生物学的サンプルから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においてハイブリダイゼーションを検出する。或いは、レーザ脱着及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及び相対存在度は、算定し得る。一実施例におけるマイクロアレイの調整及び使用について、以下に詳述する。

20

30

【0251】

組織または細胞サンプルの準備

グアニジウムチオシアネート法を用いて組織サンプルから全RNAを単離し、オリゴ（dT）セルロース法を用いてポリ（A）⁺ RNAを精製する。各ポリ（A）⁺ RNAサンプルを、MLV逆転写酵素、0.05 pg/μlのオリゴ（dT）プライマー（21mer）、1×第一鎖合成バッファー、0.03 unit/μlのRNA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）を用いて逆転写する。逆転写反応は、GEMRI GHTキット（Incyte）を用いて、200 ngのポリ（A）⁺ RNAを含む25 ml容量で行う。特異的対照ポリ（A）⁺ RNAは、非コード酵母ゲノムDNAからin vitro転写により合成する。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。サンプルは、2つの連続するCHROMASPIN 30ゲル濾過スピナラム（CLONTECH Laboratories, Inc.（CLONTECH）, Palo Alto CA）を用いて精製する。結合後、2つの反応サンプルは、1 mlのグリコーゲン（1 mg/ml）を用いて析出させたエタノール、60 mlの酢酸ナトリウム及び300 mlの100%エタノールである。

40

50

サンプルは次に、Speed VAC (Savant Instruments Inc., Holbrook NY) を用いて乾燥して仕上げ、 $14 \mu\text{l}$ $5 \times \text{SSC} / 0.2\% \text{SDS}$ 中で再懸濁する。

【0252】

マイクロアレイの準備

本発明の配列を用いて、アレイエレメントを生成する。各アレイエレメントは、クローン化 cDNA インサートによりベクター含有細菌性細胞から増幅する。PCR 増幅は、cDNA インサートの側面に位置するベクター配列に相補的なプライマーを用いる。30 サイクルの PCR で $1 \sim 2 \text{ ng}$ の初期量から $5 \mu\text{g}$ より大きい最終量までアレイエレメントを増幅する。増幅されたアレイエレメントは、SEPHACRYL - 400 (Amersham Pharmacia Biotech) を用いて精製される。

10

【0253】

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、処理中及び処理後に大量の蒸留水洗液を用いて 0.1% の SDS 及びアセトン中で超音波により洗浄する。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、蒸留水中で広範囲にわたって洗浄し、 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ガラスは、 110°C の天火で硬化させる。

20

【0254】

米国特許第 5,807,522 号で説明されている方法を用いて、コーティングしたガラス基板にアレイエレメントを付加する。該特許は、引用を以って本明細書の一部となす。平均濃度が $100 \text{ ng} / \mu\text{l}$ のアレイエレメント DNA $1 \mu\text{l}$ を高速ロボット装置により開口キャピラリープリントエレメントに充填する。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加える。

【0255】

マイクロアレイには、STRATALINKER UV 架橋剤 (Stratagene) を用いて UV 架橋する。マイクロアレイは、室温において $0.2\% \text{SDS}$ で 1 度洗浄し、蒸留水で 3 度洗浄する。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°C で 30 分間マイクロアレイをインキュベートした後、前に行ったように $0.2\% \text{SDS}$ 及び蒸留水で洗浄することにより、非特異結合部位をブロックする。

30

【0256】

ハイブリダイゼーション

ハイブリダイゼーション反応は、 $5 \times \text{SSC}$, $0.2\% \text{SDS}$ ハイブリダイゼーション緩衝液中の Cy 3 及び Cy 5 標識した cDNA 合成生成物を各 $0.2 \mu\text{g}$ 含む $9 \mu\text{l}$ のサンプル混合体を有する。サンプル混合体は、 65°C まで 5 分間加熱し、マイクロアレイ表面上で等分して 1.8 cm^2 のカバーガラスで覆う。アレイは、顕微鏡スライドより僅かに大きいキャビティを有する防水チェンバーに移行させる。チェンバーのコーナーに $140 \mu\text{l}$ の $5 \times \text{SSC}$ を加えることにより、チェンバー内部を湿度 100% に保持する。アレイを含むチェンバーは、 60°C 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 ($1 \times \text{SSC}$, $0.1\% \text{SDS}$) において 45°C で 10 分間洗浄し、第 2 洗浄緩衝液中 ($0.1 \times \text{SSC}$) において 45°C で 10 分間各々 3 度洗浄して乾燥させる。

40

【0257】

検出

レポーター標識ハイブリダイゼーション複合体は、Cy 3 の励起のためには 488 nm 、Cy 5 の励起のためには 632 nm でスペクトル線を生成し得る Innova 70 混合ガス 10 W レーザ (Coherent, Inc., Santa Clara CA) を備えた顕微鏡で検出する。 $20 \times$ 顕微鏡対物レンズ (Nikon, Inc., Melv

50

ille NY)を用いて、アレイ上に励起レーザー光を集中させる。アレイを含むスライドを顕微鏡のコンピュータ制御X-Yステージに置き、対物レンズを通過してラスタスキャンする。本実施例で用いた1.8cm×1.8cmのアレイは、20μmの解像度でスキャンした。

【0258】

2つの異なるスキャンのうち、混合ガスマルチラインレーザは2つの蛍光色素を連続的に励起する。放射された光は、2つの蛍光色素に応じて波長に基づき2つの光電子増倍管検出器(PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ)に分割される。アレイと光電子増倍管間に設置された好適なフィルターを用いて、シグナルをフィルタリングする。用いる蛍光色素の最大発光は、Cy3では565nm、Cy5では650nmである。装置は両方の蛍光色素からのスペクトルを同時に記録し得るが、レーザ源において好適なフィルターを用いて各アレイを通常2度スキャンし、蛍光色素1つにつき1度スキャンする。

10

【0259】

スキャンの感度は通常、既知濃度のサンプル混合体に添加されるcDNA対照種により生成されるシグナル強度を用いて校正する。アレイ上の特定の位置には相補的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ジング種の重量比1:100,000に相関させる。異なる源(例えば試験及び対照細胞を表す)からの2つのサンプルを、各々異なる蛍光色素で標識し、他と異なって発現した遺伝子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、2つの蛍光色素を有する校正cDNAの標識サンプルにより校正し、ハイブリダイゼーション混合体に各々等量を加える。

20

【0260】

光電子増倍管の出力は、IBMコンパチブルPCコンピュータにインストールされた12ビットRTI-835Hアナログ-デジタル(AID)変換ボード(Analog Devices, Inc., Norwood MA)を用いてデジタル化される。デジタル化されたデータは、青色(低シグナル)から赤色(高シグナル)までの擬似カラー範囲へのリニア20色変換を用いてシグナル強度がマッピングされたような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つの異なる蛍光色素を同時に励起及び測定する場合には、各蛍光色素の発光スペクトルを用いて、データは先ず蛍光色素と蛍光色素の間の光学磁気プリンティング(発光スペクトルの重畳に起因する)に集められる。

30

【0261】

グリッドは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、GEMTOOLS遺伝子発現分析プログラム(Incyte)である。

【0262】

1.1 相補的ポリヌクレオチド

ATRS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、天然のATRSの発現を低下させるため即ち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、これより小さな或いは大きな配列の断片の場合でも本質的に同じ方法を用いることができる。Oligo4.06ソフトウェア(National Biosciences)及びATRSのコーディング配列を用いて、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阻害する。翻訳を阻害するためには、相補的なオリゴヌクレオチドを設計して、リボソームがATRSをコードする転写物に結合するのを阻害する。

40

【0263】

1.2 ATRSの発現

ATRSの発現及び精製は、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

50

できる。細菌で A T R S が発現するために、抗生物質耐性及び c D N A の転写レベルを高める誘導性のプロモーターを含む好適なベクターに c D N A をサブクローニングする。このようなプロモーターには、l a c オペレーター調節エレメントに関連する T 5 または T 7 バクテリオファージプロモーター及び t r p - l a c (t a c) 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、B L 2 1 (D E 3) 等の好適な細菌宿主に形質転換する。抗生物質耐性をもつ細菌が、イソプロピル - D チオガラクトピラノシド (I P T G) で誘発されると A T R S を発現する。真核細胞での A T R S 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ている A u t o g r a p h i c a c a l i f o r n i c a 核多角体病ウイルス (A c M N P V) を感染させて行う。バキュロウイルスの非必須の多角体遺伝子を、相同組換え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のどちらかによって、A T R S をコードする c D N A と置換する。ウイルスの感染力は維持され、強いポリヘドリンプロモーターによって高いレベルの c D N A の転写が行われる。組換えバキュロウイルスは、多くの場合は S p o d o p t e r a f r u g i p e r d a (S f 9) 昆虫細胞に感染に用いられるが、ヒト肝細胞の感染にも用いられることもある。後者の感染の場合は、バキュロウイルスの更なる遺伝的変更が必要になる。(E n g e l h a r d , E . K . ら (1 9 9 4) P r o c . N a t l . A c a d . S c i . U S A 9 1 : 3 2 2 4 - 3 2 2 7 , S a n d i g , V . ら (1 9 9 6) H u m . G e n e T h e r . 7 : 1 9 3 7 - 1 9 4 5 . 等を参照) 。

10

20

【 0 2 6 4 】

殆どの発現系では、A T R S が、例えばグルタチオン S トランスフェラーゼ (G S T) 、または F L A G や 6 - H i s 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が素早く 1 回で行うことができる。G S T は日本住血吸虫からの 2 6 k D a の酵素であり、タンパク質の活性及び抗原性を維持した状態で、固定化グルタチオン上で融合タンパク質の精製を可能とする (A m e r s h a m P h a r m a c i a B i o t e c h) 。精製の後、G S T 部分を特定の操作部位で A T R S からタンパク質分解的に切断できる。F L A G 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及びポリクローナル抗 F L A G 抗体 (E a s t m a n K o d a k) を用いて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 - H i s は、金属キレート樹脂 (Q I A G E N) 上での精製を可能にする。タンパク質の発現及び精製の方法は、前出の A u s u b e l (1 9 9 5) 1 0 章、1 6 章に記載されている。これらの方法で精製した A T R S を直接用いて以下の実施例 1 6 、 1 7 、及び 1 8 のアッセイを行うことができる。

30

40

50

【 0 2 6 5 】

1 3 機能的アッセイ

A T R S 機能は、哺乳動物細胞培養系において生理学的に高められたレベルでの A T R S をコードする配列の発現によって評価する。c D N A を高いレベルで発現する強いプロモーターを含む哺乳動物発現ベクターに c D N A をサブクローニングする。選り抜きのベクターには、p C M V S P O R T プラスミド (L i f e T e c h n o l o g i e s) 及び p C R 3 . 1 プラスミド (I n v i t r o g e n) が含まれ、どちらもサイトメガロウイルスプロモーターを有する。リポソーム製剤或いは電気穿孔法を用いて、5 ~ 1 0 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時的に形質移入する。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、c D N A の組換えベクターからの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 (G F P ; C l o n t e c h) 、C D 6 4 または C D 6 4 - G F P 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 (F C M) を用いて、G F P または C D 6 4 - G F P を発現する形質移入された細胞を同定し、その細胞のアポトーシス状態や他の細胞特性を評価する。F C M は、細胞死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、プロピジウムヨウ化物によるDNA染色によって計測される核DNA内容物の変化、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及び細胞内におけるタンパク質の発現の変化、及び蛍光複合アネキシン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化とがある。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。

【0266】

遺伝子発現におけるATRSの影響は、ATRS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、形質転換された細胞表面で発現し、ヒト免疫グロブリンG(IgG)の保存領域と結合する。形質転換された細胞と形質転換されない細胞とは、ヒトIgGがCD64に対する抗体のどちらかで被覆された磁気ビードを用いて分離することができる(DYNAL, Lake Success, NY)。mRNAは、当分野で周知の方法で細胞から精製することができる。ATRS及び目的の他の遺伝子をコードするmRNAの発現は、ノーザン分析やマイクロアレイ技術で分析することができる。

【0267】

1.4 ATRSに特異的な抗体の作製

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実質的に精製されたATRSを用いて、標準的なプロトコルでウサギを免疫化して抗体を作り出す。

【0268】

或いは、レーザGENEソフトウェア(DNASTAR)を用いてATRSアミノ酸配列を解析し、免疫抗原性の高い領域を決定する。そして対応するオリゴペプチドを合成し、このオリゴペプチドを用いて当業者によく知られている方法で抗体を生成する。例えばC末端付近或いは隣接する親水性領域等の、適切なエピトープの選択については、当分野で公知である(前出のAusubel, 1995, 11章等を参照)。

【0269】

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、免疫抗原性を高める(前出のAusubel, 1995等を参照)。完全フロイントアジュバントにおいてオリゴペプチド-KLM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及び抗ATRS活性を検査するには、ペプチドまたはATRS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されたヤギ抗ウサギIgGと反応させる。

【0270】

1.5 特異的な抗体を用いる天然ATRSの精製

天然ATRS或いは組換えATRSを、ATRS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)のような活性化クロマトグラフィー用レジンと抗ATRS抗体とを共有結合させることにより形成する。結合後に、製造者の使用説明書に従って樹脂をブロックし、洗浄する。

【0271】

ATRSを含む培養液をイムノアフィニティークラムに通し、ATRS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A T R Sとの結合を切るような条件で（例えば、p H 2 ~ 3のバッファー、或いは高濃度の尿素またはチオシアン酸塩イオンのようなカオトロピックイオンで）溶出させ、A T R Sを回収する。

【0272】

1.6 A T R Sと相互作用する分子の同定

A T R Sまたは生物学的に活性であるA T R S断片を、 125 I ボルトンハンター試薬で標識する。（B o l t o n A . E . 及び W . M . H u n t e r （1973）B i o c h e m . J . 133：529 - 539を参照）。マルチウェルプレートに予め配列しておいた候補の分子を、標識したA T R Sと共にインキュベートし、洗浄して、標識したA T R S複合体を有する全てのウェルをアッセイする。様々なA T R S濃度で得られたデータをを用いて、候補分子と結合したA T R Sの数量及び親和性、会合についての値を計算する。

10

【0273】

別法では、A T R Sと相互作用する分子を、F i e l d s , S . 及びO . S o n g （1989, N a t u r e 340：245 - 246）に記載の酵母2 - ハイブリッドシステム（y e a s t t w o - h y b r i d s y s t e m）やM A T C H M A K E Rシステム（C l o n t e c h）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。

【0274】

A T R Sはまた、高処理型の酵母2ハイブリッドシステムを使用するP A T H C A L L I N Gプロセス（C u r a G e n C o r p . , N e w H a v e n C T）に用いて、遺伝子の2つの大きな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決定することができる（N a n d a b a l a n , K . 他（2000）米国特許第6,057,101号）。

20

【0275】

1.7 A T R S活性の実証

t R N A合成酵素活性は、 $[^{14}C]$ - 標識されたアミノ酸の存在で t R N A 基質のアミノアシル化として測定される。A T R Sは緩衝液で、 $[^{14}C]$ - 標識されたアミノ酸と好適な同族の t R N A（例えば、 $[^{14}C]$ アラニンおよび t R N A^{a1a}）とともにインキュベートされる。 ^{14}C - 標識の産物は遊離 $[^{14}C]$ アミノ酸からクロマトグラフィーによって分離され、組み込まれた $[^{14}C]$ はシンチレーションカウンタで測定される。 ^{14}C - 標識の産物の量は、このアッセイのA T R Sの活性に比例する。

30

【0276】

1.8 A T R Sアゴニスト及びアンタゴニストの同定

A T R S活性化若しくは抑制のアゴニスト若しくはアンタゴニストは、実施例17で記述したアッセイを用いてテストしても良い。アゴニストは、A T R S活性の増加を引き起こし、アンタゴニストはA T R S活性の減少を引き起こす。

【0277】

当業者は、本発明の範囲及び精神から逸脱することなく本発明の記載した方法及びシステムの種々の改変を行い得る。本発明について説明するにあたり特定の好適実施例に関連して説明を行ったが、本発明の範囲が、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。実際に、分子生物学または関連分野の専門家には明らかな、本明細書に記載されている本発明の実施方法の様々な改変は、特許請求の範囲内にあるものとする。

40

【0278】

（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及びポリペプチド配列の命名法の概略を示す。

【0279】

50

表 2 は、G e n B a n k 識別番号及び本発明のポリペプチドに最も近い G e n B a n k 相同体の注釈を示す。各ポリペプチドとその G e n B a n k 相同体が一致する確率スコアも併せて示す。

【 0 2 8 0 】

表 3 は、予測されるモチーフ及びドメインを含む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ための方法、アルゴリズム及び検索可能なデータベースと共に示す。

【 0 2 8 1 】

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 c D N A 及び / またはゲノム D N A 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。

10

【 0 2 8 2 】

表 5 は、本発明のポリヌクレオチドの代表的な c D N A ライブラリを示す。

【 0 2 8 3 】

表 6 は、表 5 に示した c D N A ライブラリの作製に用いた組織及びベクターを説明する付表である。

【 0 2 8 4 】

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、引用文献及び閾値パラメータと共に示す。

【 表 1 】

表 1

Incyte プロジ エクト 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
4574912	1	4574912CD1	5	4574912CB1
7475765	2	7475765CD1	6	7475765CB1
7475776	3	7475776CD1	7	7475776CB1
5332221	4	5332221CD1	8	5332221CB1

10

20

30

40

表 2

ポリペプチド SEQ ID NO:	Incyte ポリペプチド ID	GenBank ID NO:	塩基スコア	GenBank 相同体
1	4574912CD1	95458823	1.4e-81	システイニル tRNA 合成酵素 (cysS) [Pyrococcus abyssi]
2	7475765CD1	91001357	7.8e-104	アスパラギニル tRNA 合成酵素 [ラン藻] Kaneko, T. 他 (1995) Sequence analysis of the genome of the unicellular cyanobacterium <i>Synechocystis</i> sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome. <i>DNA Res</i> 2: 153-166; 191-198; Kaneko, T. 他 (1996) Sequence analysis of the genome of the unicellular cyanobacterium <i>Synechocystis</i> sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. <i>DNA Res</i> 3: 109-136.
3	7475776CD1	931545	3.6e-138	バリン tRNA 合成酵素 [ヒト] Hsieh, S. L. 及び Campbell, R. D. (1991) Evidence that gene G7a in the human major histocompatibility complex encodes valyl-tRNA synthetase. <i>Biochem. J.</i> 278: 809-816; Erratum in: (1992) <i>Biochem. J.</i> 281: 879.
4	5332221CD1	9143793	2.3e-72	チロシル tRNA 合成酵素 [Bacillus caldotenax]. Jones, M. D. 他 (1986) Natural variation of tyrosyl-tRNA synthetase and comparison with engineered mutants. <i>Biochemistry</i> 22: 1887-1891.

10

20

30

40

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、モチーフおよびドメイン:	分析方法及びデータベース
1	4574912CD1	564	S141 S174 S199 S221 S267 S321 S348 S351 S372 S389 S401 S437 S455 S548 T119 T188 T327 T4 T402 T416 T46 T528 T59 T81		tRNA 合成酵素クラス I (C) ドメイン: P64-I538 アミノアシル転移 RNA シグネチャ BL00178; V82-A91; K314-N324 システイニル tRNA 合成酵素 シグネチャ PR0983: W75-A86; I112-V121; E239-C257; D270-E291 60 tRNA; システイニル; 合成酵素; システイン: DM01764 Q09860 53-612; A86-E131; I139-A503	HMMER_PFAM BLIMPS_BLOCKS BLIMPS_PRINTS BLAST_DOMO
2	7475765CD1	477	S10 S103 S200 S249 S25 S264 S268 S29 S322 S52 S55 S68 S79 S90 T184 T241 T349 T372	N186 N336 N38	tRNA 合成酵素 クラス II (D, K および N) ドメイン: P135-H473	HMMER_PFAM

10

20

30

40

表3-2

アミノアシル転移 RNA 合成酵素 クラス II シグネチャ: Q222-M285	PROFLESCAN
tRNA 合成酵素 クラス II (D, K および N) シグネチャ PF00152: I44-D66; R159-I183; S220-F256; Y431-P469	BLINPS_PFAM
合成酵素 アミノアシル tRNA リガーゼ たんぱく質 生合成 ATP 結合 アスパラギン酸 tRNA アスパルチル tRNA ASPRS リシル tRNA PD000871: R148-R418	BLAST_PRODOR
アミノアシル-転移 RNA 合成酵素 クラス II DW00328 P52276 70-512: I44-P472	BLAST_DOMO

10

20

30

40

表3-3

					AA-tRNA リガーゼ II モチーフ: F242-E260	MOTIFS
3	7475776CD1	621	S113 S171 S218 S232 S293 S299 S334 S390 S402 S462 S577 S597 S9 T125 T175 T572 T575 T583 Y32		tRNA 合成酵素 クラス I (L, L, M および V) ドメ イン; M1-E352 (スコア = -5.7; E-値 = 1.5e-15) アミノアシル転移 RNA 合 成酵素 シグネチャ BL00178: Q213-N223	HMMER_PFAM
					バリン-tRNA 合成酵素 シ グネチャ PR00986: R25-W38; D137- P158; Y168-R186	BLIMPS_BLOCKS
					合成酵素 アミノアシル tRNA たんぱく質 リガー ゼ 生合成 ATP 結合 バリ ル tRNA バリン tRNA VALRS イソロイシル tRNA PD000476: P151-R484	BLIMPS_PRINTS
						BLAST_PRODOR

10

20

30

40

表3-4

4	5332221CD1	477	T57, T116, S123, T134, T169, S215, S253, T262, S298, S365, T370, S376, S385, S408, T412, T442	N19	アミノアシル転移 RNA 合 成酵素 クラス I DM00514 P26640 506- 1094: S24-P444	BLAST-DOMO
					合成酵素 アミノアシル tRNA リガーゼ たんぱく 質 生合成 ATP 結合 チロ シル tRNA/トリプトファ ニル-tRNA 合成酵 素:PD001451:E60-R467	BLAST-PRODOM
					チロシン tRNA リガーゼ DM01240 P00952 1- 309:L38-E361	BLAST-DOMO
					アミノ酸 tRNA リガー ゼ:P82-L92	MOTIFS
					tRNA 合成酵素 クラス I (Trp および Tyr) (tRNA-synt_1b):I76- D312 アミノアシル-転移 RNA 合成酵 素:BL00178A:T83-L92; T278-N288	HMMER-PFAM BLIMPS-BLOCKS

10

20

30

40

表3-5

						アミノアシル-転移 RNA 合成酵素 クラス I シグネ チャ {aa_trna_ligase_i.p rf}:D66-G113	PROFILES-SCAN
						チロシル-TRNA 合成酵 素:PR01040A:S86- V108; PR01040B:G213-D228; PR01040C:Q234-E256; PR01040D:F267-G279	BLIMPS-PRINTS

10

20

30

40

表 4

ポリヌクレオチド SEQ ID NO:	Incyte ポリヌク レオチド ID	配列長	選択された 断片	配列断片	5'位置	3'位置
5	4574912CBI	1920	1-84 752-1103	2700694F6 (OVARUTU10)	703	1226
				7176695H1 (BRSTTMC01)	87	544
				96451182	1	419
				1511674T6 (LUNGNOT14)	1209	1891
				1437821T6 (PANCNOT08)	1218	1895
6	7475765CBI	2480	1-1453	452757F1 (TLYNNOT02)	1289	1920
				1437821F6 (PANCNOT08)	483	1099
				70997854V1	1513	2189
				8024941J2	252	800
				70995240V1	1418	2021
7	7475776CBI	2714	1-1807	70996436V1	668	1268
				7739120H1	45	484
				6323951H1 (LUNGDI02)	1	151
				70998726V1	2059	2480
				70997215V1	897	1434
8	5332221CBI	1672	1-862	71520981V1	707	1237
				3534850F6 (KIDNNOT25)	477	1145
				8065585J1	1	708
				71423933V1	1893	2511
				6550760H1 (BRAFNOT02)	1917	2663
				5974444H1 (BRAFNOT01)	1154	1839
				1680426F6 (STOMET01)	2395	2714
				71426048V1	1228	1925
				789035R6 (PROSTUT03)	700	1259
				646929R6 (BRSTTUT02)	1269	1672
				1712009X14C1	1	618
				(PROSNOT16)		
				2841527H1 (DRGNOT01)	690	955
				3222761R6 (COLNNOT03)	976	1484
				1712009X16C1	288	951
				(PROSNOT16)		

10

20

30

40

10

表 5

ポリヌクレオチドSEQ ID NO:	Incyte プロジェクト ID	代表的ライブラリ
5	4574912CB1	LUNGTUT03
6	7475765CB1	XERANOT01
7	7475776CB1	INONOT03
8	5332221CB1	COLNNOT11

20

30

40

表 6

ライブラリ	ベクター	ライブラリの説明
LUNGTTUT03	PSPORT1	ライブラリは 69 才の白人男性の肺の部分切除中に左下肺葉から採取した肺腫瘍組織から単離した RNA を用いて作製した。病理学の示すところではグレード 3 の扁平上皮細胞の侵襲性のある後遺病である。患者の病歴には急性心筋梗塞、前立腺肥大、悪性皮膚腫瘍および喫煙癖がある。
KERANOT01	PBLUESCRIPT	ライブラリは自発的に流産した黒人新生児の下肢の皮膚から取得した新生児のメラニン生成細胞から単離した RNA を用いて作製した。
LNODNOT03	PINCY	ライブラリは、気管支鏡法による肺の部分除去中の 67 才白人男性から取り除かれたリンパ腺組織から単離された RNA を用いて作製された。顕微鏡検査では、この組織は 10% の生存腫瘍のある強度な壊死であることが発見された。関連する腫瘍組織の病理学検査では、グレード 3 から 4 の扁平上皮細胞の侵襲性癌が見られた。患者の病歴には血管腫が含まれる。家族歴には、アテローム硬化型冠動脈疾患患者、良性の高血圧、鬱血性心不全症、およびアテローム硬化型冠動脈疾患が含まれる。
COLNNOT11	PSPORT1	ライブラリは 60 才の白人男性の半結腸切除時に採取した結腸組織から単離した RNA を用いて作製した。

10

20

30

40

表7-1

プログラム	説明	引用文献	パラメータ関連値
ABI FACTURA	ベクター配列を除き、核酸配列のまぎらわしい塩基を遮蔽するプログラム	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	アミノ酸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。	Applied Biosystems, Foster City, CA. Paracel Inc., Pasadena, CA.	不一致<50%
ABI オートアセンブラ	核酸配列の構築を行うプログラム。	Applied Biosystems, Foster City, CA.	
BLAST	アミノ酸および核酸配列の配列類似性検索に有用なバージョン-カルアラインメント検索ツール。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) Nucleic Acids Res. 25:3389-3402.	ESTs: 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は最低5つの機能からなる: fasta, tfasta, fastx, ifastxおよびkssearch。	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; およびSmith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E 値=1.0E-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BLOKS IMPROVED サーチャー。	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; およびAttwood, T.K. 他。 (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	確率値=1.0E-3以下
HMIMER	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たマルコフモデル(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; Somnhammer, E.L.L. 他 (1988) Nucleic Acids Res. 26:320-322; Durbin, R. 他。 (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	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

10

20

30

40

表7-2

説明	引用文献	パラメータ関連値
ProfileScan	Gribskov, M. 他 (1988) CABIOS 4:61-66; Gribskov, M. 他 (1989) Methods Enzymol.183:146-159; Bairoch, A. 他。 Nucleic Acids Res. 25:4323-43 30) 。 25:217-221.	標準化された質のスコア≧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	Ewing, B. 他 (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	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. Gordon, D. 他 (1998) Genome Res. 8:195-202.	スコア=120以上 一致した長さ=56以上
Consed	Phrap 構築の表示、編集用グラフィカルツール。	
SPScan	タンパク質配列を走査し、分泌シグナルペプチドの存在を調べ る重量マトリックス分析プログラム。	スコア=3.5以上
TMAP	重量マトリックスを使ってタンパク質配列上の膜貫通部分を描 写し、定位(方向)を決定するプログラム。	
TMHMMER	隠れたMarkov モデル(HMM) を使ってタンパク質配列上の 膜貫通部分を推写し、定位(方向)を決定するプログラム。	
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 Nucleic Acids Res. 25:4323-43 30) 。 25:217-221; Wisconsin Package Program Manual, version 9, page M51-58, Genetics Computer Group, Madison, WI	

10

20

30

40

【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
29 November 2001 (29.11.2001)

PCT

(10) International Publication Number
WO 01/90330 A2

(51) International Patent Classification: C12N 9/00

[US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). **BANDMAN, Olga** [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). **LU, Dyung, Alhua, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US).

(21) International Application Number: PCT/US01/16808

(22) International Filing Date: 22 May 2001 (22.05.2001)

(25) Filing Language: English

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

(26) Publication Language: English

(30) Priority Data:
60/207,248 25 May 2000 (25.05.2000) US
60/208,791 1 June 2000 (01.06.2000) US
60/210,585 8 June 2000 (08.06.2000) 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, 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, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.(71) Applicant (*for all designated States except US*): **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, 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, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **TANG, Tom, Y.** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **PATTERSON, Chandra** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). **GANDHI, Ameena, R.** [US/US]; 837 Roble Avenue #1, Menlo Park, CA 94025 (US). **TRIBOULEY, Catherine, M.** [FR/US]; 1121 Tennessee Street #5, San Francisco, CA 94107 (US). **LEE, Ernestine, A.** [US/US]; 624 Kains Street, Albany, CA 94706 (US). **YAO, Monique, G.**

Published:

— without international search report and to be republished upon receipt of that 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.



WO 01/90330 A2

(54) Title: AMINOACYL tRNA SYNTHETASES

(57) Abstract: The invention provides human aminoacyl tRNA synthetases (ATRS) and polynucleotides which identify and encode ATRS. 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 ATRS.

WO 01/90330

PCT/US01/16808

AMINOACYL tRNA SYNTHETASES

TECHNICAL FIELD

5 This invention relates to nucleic acid and amino acid sequences of aminoacyl tRNA synthetases and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative and autoimmune/inflammatory disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of aminoacyl tRNA synthetases.

10 BACKGROUND OF THE INVENTION

Correct translation of the genetic code depends upon each amino acid forming a linkage with the appropriate transfer RNA (tRNA). The aminoacyl-tRNA synthetases (aaRSs) are essential proteins found in all living organisms. The aaRSs are responsible for the activation and correct attachment of an amino acid with its cognate tRNA, as the first step in protein biosynthesis.

15 Prokaryotic organisms have at least twenty different types of aaRSs, one for each different amino acid, while eukaryotes usually have two aaRSs, a cytosolic form and a mitochondrial form, for each different amino acid. The 20 aaRS enzymes can be divided into two structural classes. Class I enzymes add amino acids to the 2' hydroxyl at the 3' end of tRNAs while Class II enzymes add amino acids to the 3' hydroxyl at the 3' end of tRNAs. Each class is characterized by a distinctive topology of the catalytic domain. Class I enzymes contain a catalytic domain based on the nucleotide-binding 'Rossmann fold'. In particular, a consensus tetrapeptide motif is highly conserved (Prosite Document PDOC00161, Aminoacyl-transfer RNA synthetases class-I signature). Class I enzymes are specific for arginine, cysteine, glutamic acid, glutamine, isoleucine, leucine, methionine, tyrosine, tryptophan, and valine. Class II enzymes contain a central catalytic domain, which consists of a seven-stranded antiparallel β -sheet domain, as well as N- and C-terminal regulatory domains. Class II enzymes are separated into two groups based on the heterodimeric or homodimeric structure of the enzyme; the latter group is further subdivided by the structure of the N- and C-terminal regulatory domains (Hartlein, M. and Cusack, S. (1995) J. Mol. Evol. 40:519-530). Class II enzymes are specific for alanine, asparagine, aspartic acid, glycine, histidine, lysine, phenylalanine, proline, serine, and threonine.

25 Certain aaRSs also have editing functions. IleRS, for example, can misactivate valine to form Val-tRNA^{Ile}, but this product is cleared by a hydrolytic activity that destroys the mischarged product. This editing activity is located within a second catalytic site found in the connective polypeptide 1 region (CPI), a long insertion sequence within the Rossmann fold domain of Class I enzymes

30

WO 01/90330

PCT/US01/16808

(Schimmel, P. et al. (1998) FASEB J. 12:1599-1609). AARSs also play a role in tRNA processing. It has been shown that mature tRNAs are charged with their respective amino acids in the nucleus before export to the cytoplasm, and charging may serve as a quality control mechanism to insure the tRNAs are functional (Martinis, S.A. et al. (1999) EMBO J. 18:4591-4596).

5 In addition to their function in protein synthesis, specific aminoacyl tRNA synthetases also play roles in cellular fidelity, RNA splicing, RNA trafficking, apoptosis, and transcriptional and translational regulation. For example, human tyrosyl-tRNA synthetase can be proteolytically cleaved into two fragments with distinct cytokine activities. The carboxy-terminal domain exhibits monocyte and leukocyte chemotaxis activity as well as stimulating production of myeloperoxidase, tumor
10 necrosis factor- α , and tissue factor. The N-terminal domain binds to the interleukin-8 type A receptor and functions as an interleukin-8-like cytokine. Human tyrosyl-tRNA synthetase is secreted from apoptotic tumor cells and may accelerate apoptosis (Wakasugi, K., and Schimmel, P. (1999) Science 284:147-151). Mitochondrial *Neurospora crassa* TyrRS and *S. cerevisiae* LeuRS are essential factors for certain group I intron splicing activities, and human mitochondrial LeuRS can substitute for the
15 yeast LeuRS in a yeast null strain. Certain bacterial aaRSs are involved in regulating their own transcription or translation (Martinis, *supra*). Several aaRSs are able to synthesize diadenosine oligophosphates, a class of signalling molecules with roles in cell proliferation, differentiation, and apoptosis (Kisselev, L.L. et al. (1998) FEBS Lett. 427:157-163; Vartanian, A. et al. (1999) FEBS Lett. 456:175-180).

20 Autoantibodies against aminoacyl-tRNAs are generated by patients with autoimmune diseases such as rheumatic arthritis, dermatomyositis and polymyositis, and correlate strongly with complicating interstitial lung disease (ILD) (Freist, W. et al. (1999) Biol. Chem. 380:623-646; Freist, W. et al. (1996) Biol. Chem. Hoppe Seyler 377:343-356). These antibodies appear to be generated in response to viral infection, and coxsackie virus has been used to induce experimental viral myositis in animals.

25 Comparison of aaRS structures between humans and pathogens has been useful in the design of novel antibiotics (Schimmel, *supra*). Genetically engineered aaRSs have been utilized to allow site-specific incorporation of unnatural amino acids into proteins *in vivo* (Liu, D.R. et al. (1997) Proc. Natl. Acad. Sci. USA 94:10092-10097).

30 The discovery of new aminoacyl tRNA synthetases 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 and autoimmune/inflammatory disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of aminoacyl tRNA synthetases

WO 01/90330

PCT/US01/16808

SUMMARY OF THE INVENTION

The invention features purified polypeptides, aminoacyl tRNA synthetases, referred to collectively as "ATRS" and individually as "ATRS-1," "ATRS-2," "ATRS-3," and "ATRS-4". 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:1-4, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-4.

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:1-4, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-4. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:5-8.

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:1-4, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4. 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.

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

WO 01/90330

PCT/US01/16808

of SEQ ID NO:1-4, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4. 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 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:1-4, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4.

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:5-8, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:5-8, 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:5-8, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:5-8, 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

WO 01/90330

PCT/US01/16808

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:5-8, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:5-8, 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:1-4, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-4. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional ATRs, 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:1-4, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4. 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

WO 01/90330

PCT/US01/16808

pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional ATRS, comprising administering to a patient in need of such treatment the composition.

5 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:1-4, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, and
10 d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4. 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
15 treating a disease or condition associated with overexpression of functional ATRS, 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:1-4, b) a naturally occurring polypeptide
20 comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4. The method comprises a) combining the polypeptide with at least one test compound
25 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:1-4, b) a naturally occurring
30 polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group

WO 01/90330

PCT/US01/16808

consisting of SEQ ID NO:1-4. 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.

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 sequence selected from the group consisting of SEQ ID NO:5-8, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

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) 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:5-8, ii) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:5-8, iii) a polynucleotide 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:5-8, ii) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:5-8, 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 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.

WO 01/90330

PCT/US01/16808

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

5 Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is 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.

10 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.

15 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.

DESCRIPTION OF THE INVENTION

20 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.

25 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.

30 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

WO 01/90330

PCT/US01/16808

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 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.

5 DEFINITIONS

"ATRS" refers to the amino acid sequences of substantially purified ATRS obtained from 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 ATRS. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of ATRS either by directly interacting with ATRS or by acting on components of the biological pathway in which ATRS participates.

An "allelic variant" is an alternative form of the gene encoding ATRS. 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 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 ATRS include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as ATRS or a polypeptide with at least one functional characteristic of ATRS. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding ATRS, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding ATRS. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent ATRS. 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 ATRS is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and 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

WO 01/90330

PCT/US01/16808

having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

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 ATRS. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of ATRS either by directly interacting with ATRS or by acting on components of the biological pathway in which ATRS 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 ATRS 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 "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

WO 01/90330

PCT/US01/16808

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.

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 ATRS, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"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 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 ATRS or fragments of ATRS 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

WO 01/90330

PCT/US01/16808

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
5	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
10	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
15	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
20	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

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.

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 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.

"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

WO 01/90330

PCT/US01/16808

diseased and a normal sample.

A "fragment" is a unique portion of ATRS or the polynucleotide encoding ATRS which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a
 5 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 example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected
 10 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 embodiments.

A fragment of SEQ ID NO:5-8 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:5-8, for example, as distinct from any other sequence in the genome
 15 from which the fragment was obtained. A fragment of SEQ ID NO:5-8 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:5-8 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:5-8 and the region of SEQ ID NO:5-8 to which the fragment corresponds are routinely determinable by one of
 20 ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-4 is encoded by a fragment of SEQ ID NO:5-8. A fragment of SEQ ID NO:1-4 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-4. For example, a fragment of SEQ ID NO:1-4 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-4. The precise length of a
 25 fragment of SEQ ID NO:1-4 and the region of SEQ ID NO:1-4 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
 30 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

WO 01/90330

PCT/US01/16808

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.

5 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
10 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 similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms
15 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
20 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/b12.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
25 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:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

30 *Open Gap: 5 and Extension Gap: 2 penalties*

Gap x drop-off: 50

Expect: 10

Word Size: 11

WO 01/90330

PCT/US01/16808

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.

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.

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 3.12c 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*

WO 01/90330

PCT/US01/16808

*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.

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)

WO 01/90330

PCT/US01/16808

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 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 ATRS 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 ATRS 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.

WO 01/90330

PCT/US01/16808

The term "modulate" refers to a change in the activity of ATRS. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of ATRS.

5 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.

10 "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 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.

20 "Post-translational modification" of an ATRS 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 ATRS.

"Probe" refers to nucleic acid sequences encoding ATRS, 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).

30 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

WO 01/90330

PCT/US01/16808

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 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 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 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 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 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

WO 01/90330

PCT/US01/16808

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 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
 5 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.

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
 10 expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated 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.

15 "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
 20 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 ATRS, nucleic acids encoding ATRS, or fragments thereof may comprise a bodily fluid; an extract from a
 25 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
 30 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.

WO 01/90330

PCT/US01/16808

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.

5 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,
10 trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression 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
15 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
20 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
25 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. 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,
30 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.

WO 01/90330

PCT/US01/16808

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 polynucleotides due to alternative 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 aminoacyl tRNA synthetases (ATRS), the polynucleotides encoding ATRS, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative and autoimmune/inflammatory disorders.

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

WO 01/90330

PCT/US01/16808

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.

5 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 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. 10 Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog 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. 15 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 20 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 aminoacyl tRNA synthetases. For example, SEQ ID NO:1 is 41% identical from amino acid residues 51 to 503 to *Pyrococcus abyssi* cysteinyl- 25 tRNA synthetase (GenBank ID g5458823) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.4e-81, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a tRNA synthetase class I (C) 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 analyses provide further corroborative evidence that SEQ ID NO:1 is 30 a cysteinyl-tRNA synthetase. In an alternate example, SEQ ID NO:2 is 46% identical to *Synechocystis* sp. asparaginyl-tRNA synthetase (GenBank ID g1001357) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 7.8e-104,

WO 01/90330

PCT/US01/16808

which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains a tRNA synthetase class II (D, K and N) 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

5 PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:2 is an asparaginyl-tRNA synthetase. In a further example, SEQ ID NO:4 is 39% identical to Bacillus caldotenax tyrosyl-tRNA synthetase (GenBank ID g143793) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $2.3e-72$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ
10 ID NO:4 also contains an tyrosyl-tRNA synthetase 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:4 is a tRNA synthetase. SEQ ID NO:3 was analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ
15 ID NO:1-4 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide
20 consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:5-8 or that distinguish between SEQ ID NO:5-8 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA
25 sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

30 The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 2700694F6 is the identification number of an Incyte cDNA sequence, and OVARTUT10 is the cDNA library from which it is derived. Alternatively, the identification numbers in column 5 may refer to GenBank

WO 01/90330

PCT/US01/16808

cDNAs or ESTs (e.g., g6451182) which contributed to the assembly of the full length polynucleotide sequences. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 70997854V1). Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 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 ATRS variants. A preferred ATRS 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 ATRS amino acid sequence, and which contains at least one functional or structural characteristic of ATRS.

The invention also encompasses polynucleotides which encode ATRS. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:5-8, which encodes ATRS. The polynucleotide sequences of SEQ ID NO:5-8, 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.

The invention also encompasses a variant of a polynucleotide sequence encoding ATRS. In particular, such a variant polynucleotide sequence will have at least about 80%, or alternatively at least about 90%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding ATRS. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:5-8 which has at least about 70%, or alternatively at least about 85%, or even at least about 95%

WO 01/90330

PCT/US01/16808

polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:5-8. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of ATRS.

5 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 ATRS, some bearing minimal 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
10 polynucleotide sequence of naturally occurring ATRS, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode ATRS and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring ATRS under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding ATRS or
15 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 ATRS and its derivatives without altering the encoded amino acid sequences
20 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 ATRS and ATRS 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
25 using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding ATRS 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:5-8 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and
30 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

WO 01/90330

PCT/US01/16808

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 ATRS may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, 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. 16:8186.) 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.

WO 01/90330

PCT/US01/16808

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
5 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
10 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.

15 In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode ATRS may be cloned in recombinant DNA molecules that direct expression of ATRS, 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 ATRS.

20 The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter ATRS-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
25 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 MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat.
30 Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of ATRS, 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

WO 01/90330

PCT/US01/16808

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
5 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 ATRS may be synthesized, in whole or in part,
10 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, ATRS 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.,
15 Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 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 ATRS, 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
20 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.)

In order to express a biologically active ATRS, the nucleotide sequences encoding ATRS or
25 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 ATRS. Such elements may vary in their strength and specificity. Specific initiation signals
30 may also be used to achieve more efficient translation of sequences encoding ATRS. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding ATRS and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may

WO 01/90330

PCT/US01/16808

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. 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 ATRS 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 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 ATRS. These include, but are not limited to, microorganisms such as bacteria transformed 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 animal cell systems. (See, e.g., Sambrook, *supra*; Ausubel, *supra*; Van Hecke, 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 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) *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 ATRS. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding ATRS can be achieved using a

WO 01/90330

PCT/US01/16808

multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding ATRS 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

5 *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of 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 ATRS are needed, e.g. for the production of antibodies, vectors which direct high level expression of ATRS may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

10 Yeast expression systems may be used for production of ATRS. A number of vectors 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*;

15 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 ATRS. Transcription of sequences encoding ATRS 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

20 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,

25 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 ATRS 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

30 infective virus which expresses ATRS 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.

WO 01/90330

PCT/US01/16808

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 ATRS in cell lines is preferred. For example, sequences encoding ATRS 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.

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 ATRS is inserted within a marker gene sequence, transformed cells containing sequences encoding ATRS can be identified by the absence of marker gene function. Alternatively, a

WO 01/90330

PCT/US01/16808

marker gene can be placed in tandem with a sequence encoding ATRS 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 ATRS and that express ATRS 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 ATRS 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 ATRS is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, 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.)

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 ATRS include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding ATRS, or any fragments thereof, may be cloned into a vector 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 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 ATRS 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

WO 01/90330

PCT/US01/16808

and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode ATRS may be designed to contain signal sequences which direct secretion of ATRS through a prokaryotic or eukaryotic cell membrane.

5 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 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
10 (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture 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 ATRS may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric ATRS protein
15 containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of ATRS activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity 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
20 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 proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize
25 these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the ATRS encoding sequence and the heterologous protein sequence, so that ATRS 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 available kits may also be used to facilitate expression and purification of fusion proteins.

30 In a further embodiment of the invention, synthesis of radiolabeled ATRS 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

WO 01/90330

PCT/US01/16808

precursor, for example, ³⁵S-methionine.

ATRS of the present invention or fragments thereof may be used to screen for compounds that specifically bind to ATRS. At least one and up to a plurality of test compounds may be screened for specific binding to ATRS. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of ATRS, 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 ATRS binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express ATRS, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing ATRS or cell membrane fractions which contain ATRS are then contacted with a test compound and binding, stimulation, or inhibition of activity of either ATRS 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 ATRS, either in solution or affixed to a solid support, and detecting the binding of ATRS 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.

ATRS of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of ATRS. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for ATRS activity, wherein ATRS is combined with at least one test compound, and the activity of ATRS in the presence of a test compound is compared with the activity of ATRS in the absence of the test compound. A change in the activity of ATRS in the presence of the test compound is indicative of a compound that modulates the activity of ATRS. Alternatively, a test compound is combined with an in vitro or cell-free system comprising ATRS under conditions suitable for ATRS activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of ATRS may do so indirectly and need not come in direct contact with the test compound. At least one and up to a

WO 01/90330

PCT/US01/16808

plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding ATRS 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 Number 5,175,383 and U.S. Patent Number 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 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 ATRS 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 ATRS 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 ATRS 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 ATRS, e.g., by secreting ATRS 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 ATRS and aminoacyl tRNA synthetases. In addition, the expression of ATRS is closely

WO 01/90330

PCT/US01/16808

associated with lung tumor tissue, with neonatal keratinocytes and lymph node tissue, and with disease states of the colon and prostate. Therefore, ATRS appears to play a role in cell proliferative and autoimmune/inflammatory disorders. In the treatment of disorders associated with increased ATRS expression or activity, it is desirable to decrease the expression or activity of ATRS. In the treatment of disorders associated with decreased ATRS expression or activity, it is desirable to increase the expression or activity of ATRS.

Therefore, in one embodiment, ATRS 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 ATRS. 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, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and 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.

In another embodiment, a vector capable of expressing ATRS 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 ATRS including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified ATRS in

WO 01/90330

PCT/US01/16808

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 ATRS including, but not limited to, those provided above.

5 In still another embodiment, an agonist which modulates the activity of ATRS may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ATRS including, but not limited to, those listed above.

10 In a further embodiment, an antagonist of ATRS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of ATRS. Examples of such disorders include, but are not limited to, those cell proliferative and autoimmune/inflammatory disorders described above. In one aspect, an antibody which specifically binds ATRS 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 ATRS.

15 In an additional embodiment, a vector expressing the complement of the polynucleotide encoding ATRS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of ATRS including, but not limited to, those described above.

20 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 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.

25 An antagonist of ATRS may be produced using methods which are generally known in the art. In particular, purified ATRS may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind ATRS. Antibodies to ATRS 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 fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

30 For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with ATRS 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

WO 01/90330

PCT/US01/16808

such as aluminum hydroxide, and surface active substances such 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 ATRS 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 ATRS amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to ATRS 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 Cole, S.P. et al. (1984) *Mol. Cell Biol.* 6:109-120.)

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. 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 ATRS-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, 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.)

Antibody fragments which contain specific binding sites for ATRS 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.

WO 01/90330

PCT/US01/16808

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 immunoassays typically involve the measurement of complex formation between ATRS and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering ATRS epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for ATRS. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of ATRS-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 multiple ATRS epitopes, represents the average affinity, or avidity, of the antibodies for ATRS. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular ATRS 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 ATRS-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of ATRS, 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).

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 ATRS-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for 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 ATRS, 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,

WO 01/90330

PCT/US01/16808

RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding ATRS. 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 ATRS. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

5 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) 10 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.* 15 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding ATRS 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 25 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) 30 *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 ATRS expression or regulation causes disease, the expression of

WO 01/90330

PCT/US01/16808

ATRS 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 ATRS are treated by constructing mammalian expression vectors encoding ATRS and introducing these vectors by mechanical means into ATRS-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.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J.-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of ATRS include, but are not limited to, the pCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). ATRS 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 (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 PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding ATRS from a normal individual.

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. (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 ATRS expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding ATRS under the control of an independent promoter or the retrovirus long

WO 01/90330

PCT/US01/16808

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. 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 Number 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 ATRS to cells which have one or more genetic abnormalities with respect to the expression of ATRS. 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 (Csote, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinuzzi, 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 ATRS to target cells which have one or more genetic abnormalities with respect to the expression of ATRS. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing ATRS 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

WO 01/90330

PCT/US01/16808

been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant
5 HSV d92 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
10 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.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to
15 deliver polynucleotides encoding ATRS 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,
20 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 ATRS into the alphavirus genome in place of the capsid-coding region results in the production of a large number of ATRS-coding RNAs and the synthesis of high levels of ATRS in vector transduced cells. While alphavirus
25 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 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 ATRS into a variety of cell types. The specific transduction of a subset of cells in a
30 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

WO 01/90330

PCT/US01/16808

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 been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr,

5 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.

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
10 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 ATRS.

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,
15 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. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA
20 sequences encoding ATRS. 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
25 constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

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
30 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, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine,

WO 01/90330

PCT/US01/16808

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 ATRS. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not
 5 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 ATRS
 10 expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding ATRS may be therapeutically useful, and in the treatment of disorders associated with decreased ATRS expression or activity, a compound which specifically promotes expression of the polynucleotide encoding ATRS may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in
 15 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
 20 library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding ATRS 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 ATRS are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected
 25 by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding ATRS. 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
 30 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.

WO 01/90330

PCT/US01/16808

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 (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

5 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. 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.

10 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 monkeys.

15 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 Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of ATRS, antibodies to ATRS, and mimetics, agonists, antagonists, or inhibitors of ATRS.

20 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, sublingual, or rectal means.

25 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 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

30 needle injection, and obviates the need for potentially toxic penetration enhancers. Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination

WO 01/90330

PCT/US01/16808

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 ATRS or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, ATRS 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 ATRS or fragments thereof, antibodies of ATRS, and agonists, antagonists or inhibitors of ATRS, 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

WO 01/90330

PCT/US01/16808

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 inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind ATRS may be used for the diagnosis of disorders characterized by expression of ATRS, or in assays to monitor patients being treated with ATRS or agonists, antagonists, or inhibitors of ATRS. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for ATRS include methods which utilize the antibody and a label to detect ATRS in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, 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 ATRS, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of ATRS expression. Normal or standard values for ATRS expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to ATRS under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of ATRS 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 ATRS 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 ATRS may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of ATRS, and to monitor regulation of ATRS levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding ATRS or closely related molecules may be used to identify nucleic acid sequences which encode ATRS. 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 conserved motif, and the stringency of the hybridization or amplification will determine whether the

WO 01/90330

PCT/US01/16808

probe identifies only naturally occurring sequences encoding ATRS, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the ATRS 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:5-8 or from genomic sequences including promoters, enhancers, and introns of the ATRS gene.

Means for producing specific hybridization probes for DNAs encoding ATRS include the cloning of polynucleotide sequences encoding ATRS or ATRS 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, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding ATRS may be used for the diagnosis of disorders associated with expression of ATRS. 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, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and 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

WO 01/90330

PCT/US01/16808

syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma. The polynucleotide sequences encoding ATRS may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in
5 microarrays utilizing fluids or tissues from patients to detect altered ATRS expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding ATRS may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding ATRS may be labeled by standard methods and added to a fluid or tissue sample
10 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 sample then the presence of altered levels of nucleotide sequences encoding ATRS in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the
15 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 ATRS, a normal or standard profile for expression is established. This may be accomplished by combining
20 body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding ATRS, 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 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
25 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 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
30 days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or 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

WO 01/90330

PCT/US01/16808

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.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding ATRS 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 ATRS, or a fragment of a polynucleotide complementary to the polynucleotide encoding ATRS, 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 ATRS may be used to detect single nucleotide polymorphisms (SNPs). SNPs are 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 ATRS are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal 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 (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 high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of ATRS 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. 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

WO 01/90330

PCT/US01/16808

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 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 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 profile.

In another embodiment, ATRS, fragments of ATRS, or antibodies specific for ATRS 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.

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 a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 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 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.

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

WO 01/90330

PCT/US01/16808

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 (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 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 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 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

WO 01/90330

PCT/US01/16808

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 ATRS to quantify the levels of ATRS 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. 270:103-111; Mendoza, 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.

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. Seilhamer (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 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 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

WO 01/90330

PCT/US01/16808

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.

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 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 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 ATRS may be used 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

WO 01/90330

PCT/US01/16808

(OMIM) World Wide Web site. Correlation between the location of the gene encoding ATRS 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, ATRS, 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 solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between ATRS 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 ATRS, or fragments thereof, and washed. Bound ATRS is then detected by methods well known in the art. Purified ATRS 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 ATRS specifically compete with a test compound for binding ATRS. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with ATRS.

In additional embodiments, the nucleotide sequences which encode ATRS may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on

WO 01/90330

PCT/US01/16808

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. 60/207,248, U.S. Ser. No. 60/208,791, and U.S. Ser. No. 60/210,585, 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) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of 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.

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 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 SUPERSCRIP^T plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*, 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.

WO 01/90330

PCT/US01/16808

chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , 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 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 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

WO 01/90330

PCT/US01/16808

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
 5 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, and hidden Markov model (HMM)-based protein family
 10 databases such as PFAM. (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. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or
 15 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 Consed, 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 the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may
 20 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, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering,
 25 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 (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
 30 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 which are incorporated by reference herein in their entirety, and the fourth column presents, where

WO 01/90330

PCT/US01/16808

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 polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:5-8. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative aminoacyl tRNA synthetases were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri 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 aminoacyl tRNA synthetases, the encoded polypeptides were analyzed by querying against PFAM models for aminoacyl tRNA synthetases. Potential aminoacyl tRNA synthetases were also identified by homology to Incyte cDNA sequences that had been annotated as aminoacyl tRNA synthetases. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri 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

WO 01/90330

PCT/US01/16808

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 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 gbpi 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 ATRS Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:5-8 were compared with

WO 01/90330

PCT/US01/16808

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:5-8 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 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:5 was mapped to chromosome 12 within the interval from 97.1 to 116.6 centiMorgans.

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.)

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:

WO 01/90330

PCT/US01/16808

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length}(\text{Seq. 1}), \text{length}(\text{Seq. 2}) \}}$$

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 (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 ATRS 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; hematologic 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 ATRS. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of ATRS Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate

WO 01/90330

PCT/US01/16808

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 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 CviII 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

WO 01/90330

PCT/US01/16808

(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.

5 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 quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA
10 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 BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or
15 are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:5-8 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is
20 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 µCi of [γ -³²P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size
25 exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10⁷ 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
30 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

WO 01/90330

PCT/US01/16808

compared.

X. 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 (Skena (1999), supra). 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., Skena, M. et al. (1995) Science 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 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 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.

25 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/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM 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

WO 01/90330

PCT/US01/16808

with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to the 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

5 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 in 14 µl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element

10 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 amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope

15 slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water 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.

Array elements are applied to the coated glass substrate using a procedure described in US

20 Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, 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).

25 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

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and

30 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

WO 01/90330

PCT/US01/16808

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 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.

5 **Detection**

Reporter-labeled hybridization complexes are detected with a microscope equipped with an 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
10 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 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,
15 Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate 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.

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,
20 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
30 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

WO 01/90330

PCT/US01/16808

spectra) between the fluorophores using each fluorophore's emission spectrum.

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).

XI. Complementary Polynucleotides

Sequences complementary to the ATRS-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring ATRS. 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 ATRS. 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 ATRS-encoding transcript.

XII. Expression of ATRS

Expression and purification of ATRS is achieved using bacterial or virus-based expression systems. For expression of ATRS 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* (*lac*) 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 ATRS upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of ATRS in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding ATRS 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, ATRS 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,

WO 01/90330

PCT/US01/16808

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 ATRS 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 ATRS obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, where applicable.

XIII. Functional Assays

ATRS function is assessed by expressing the sequences encoding ATRS 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 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 ATRS on gene expression can be assessed using highly purified populations

WO 01/90330

PCT/US01/16808

of cells transfected with sequences encoding ATRS 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 ATRS and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of ATRS Specific Antibodies

ATRS 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 rabbits and to produce antibodies using standard protocols.

Alternatively, the ATRS 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-ATRS activity by, for example, binding the peptide or ATRS to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring ATRS Using Specific Antibodies

Naturally occurring or recombinant ATRS is substantially purified by immunoaffinity chromatography using antibodies specific for ATRS. An immunoaffinity column is constructed by covalently coupling anti-ATRS 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 ATRS are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of ATRS (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt

WO 01/90330

PCT/US01/16808

antibody/ATRS 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 ATRS is collected.

XVI. Identification of Molecules Which Interact with ATRS

ATRS, 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 ATRS, washed, and any wells with labeled ATRS complex are assayed. Data obtained using different concentrations of ATRS are used to calculate values for the number, affinity, and association of ATRS with the candidate molecules.

Alternatively, molecules interacting with ATRS 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).

ATRS 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).

XVII. Demonstration of ATRS Activity

tRNA synthetase activity is measured as the aminoacylation of a substrate tRNA in the presence of [¹⁴C]-labeled amino acid. ATRS is incubated with [¹⁴C]-labeled amino acid and the appropriate cognate tRNA (for example, [¹⁴C]alanine and tRNA^{Ala}) in a buffered solution. ¹⁴C-labeled product is separated from free [¹⁴C]amino acid by chromatography, and the incorporated ¹⁴C is quantified by scintillation counter. The amount of ¹⁴C-labeled product detected is proportional to the activity of ATRS in this assay.

XVIII. Identification of ATRS Agonists and Antagonists

Agonists or antagonists of ATRS activation or inhibition may be tested using the assay described in section XVII. Agonists cause an increase in ATRS activity and antagonists cause a decrease in ATRS 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

WO 01/90330

PCT/US01/16808

to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

WO 01/90330

PCT/US01/16808

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
4574912	1	4574912CD1	5	4574912CB1
7475765	2	7475765CD1	6	7475765CB1
7475776	3	7475776CD1	7	7475776CB1
5332221	4	5332221CD1	8	5332221CB1

WO 01/90330

PCT/US01/16808

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
1	4574912CD1	g5458823	1.4e-81	cyateinyl-tRNA synthetase (cysS) [<i>Pyrococcus abyssi</i>]
2	7475765CD1	g1001357	7.8e-104	asparaginyl-tRNA synthetase [<i>Synechocystis</i> sp.] Kaneko, T. et al. (1995) Sequence analysis of the genome of the unicellular cyanobacterium <i>Synechocystis</i> sp. strain F803. I. Genome. <i>Proc Natl Acad Sci USA</i> 92:1000-1004. Kaneko, T. et al. (1996) Sequence analysis of the genome of the unicellular cyanobacterium <i>Synechocystis</i> sp. strain F803. II. Sequence determination of the entire genome and assignment of genes to protein-coding regions. <i>DNA Res.</i> 3:109-136.
3	7475776CD1	g31545	3.6e-138	valyl-tRNA synthetase [<i>Homo sapiens</i>] Hsieh, S. L. and Campbell, R.D. (1991) Evidence that gene <i>G7a</i> in the human major histocompatibility complex encodes valyl-tRNA synthetase. <i>Biochem.</i> <i>Biophys. Res. Commun.</i> 178:187-191. Bi. <i>Cell</i> 73:281-277.
4	5332221CD1	g143793	2.3e-72	Tyrosyl-tRNA synthetase [<i>Bacillus</i> <i>Caldoferax</i>]. Jones, M.D. et al. (1986) Natural variation of tyrosyl-tRNA synthetase and comparison with engineered mutants. <i>Biochemistry</i> 22:1887-1891.

WO 01/90330

PCT/US01/16808

Table 3

					AMINOACYL-TRANSFER RNA SYNTHETASES class-II signatures: Q222-M285	PROFILESCAN
					tRNA synthetases class II (D, K and N) signature PF00152; R44-356; R159- I183; S220-P256; Y431-P469	BLIMPS_PFAM
					SYNTHETASE AMINOACYL/TRAN LIGASE PROTEIN BIOSYNTHESIS ATPBINDING ASPARTATETRNA ASPARTATETRNA ASPRAS LIGASE SYNTHETASE PF000871; R148-R418	BLAST_PRODOM
					AMINOACYL-TRANSFER RNA SYNTHETASES CLASS-II DM00328 P52276 70- 512; 144-P472	BLAST_DOMO

WO 01/90330

PCT/US01/16808

Table 3

SRQ ID	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Analytical Methods and Databases
1	4574912CD1	564	S141 S174 S199 S221 S267 S321 S346 S351 S372 S389 S401 S413 S455 S468 T119 T188 T327 T4 T402 T416 T46 T528 T59 T81		Signature: tRNA synthetases class I (C) domain: P64-I538 Aminoacyl-transfer RNA signature: EL00178 V82-A91; K314-N324 CysteinyI LRNA synthetase signature PR00983: W75-A86; I112-V121; E239-C257; V121-E239 do tRNA: CYSTEINYL; SYNTHETASE; CYSTEINE; DM01764 Q09860 53-612; A86-EL31; L139-A503 CRNA synthetases class I, O, K and N1 domain: P135-H473	HMMER_PPFAM BLIMPS_BLOCKS BLIMPS_PRINTS ELAST_DOMO HMMER_PPFAM
2	7475765CD1	477	S10 S163 S209 S249 S25 S264 S268 S29 S322 S52 S55 S68 S79 S90 T184 T241 T349 T372	M186 N336 N38		

Table 3

3	74/75776CD1	621	S113 S171 S218 S232 S293 S299 S334 S390 S402 S462 S577 S597 S9 T125 T175 Y572 Y575 Y583 Y32	aa-tRNA ligase II motif: F242-E260 tRNA synthetases class I (L, M, N and V) domain: M1-E352 (Score = -5.7; E- value = 1.5e-15) Aminoacyl-transfer RNA synthetase signature BL00178: Q213-N223 Valyl-tRNA synthetase signature PR00986: R25-W38; D137- F159; Y168-R186 SYNTHETASE aa-tRNA ligase PROTEIN LIGASE BIOSYNTHESIS ATEPENDING VALYLTRNA VALINETRNA VALRS ISOLEUCYLTRNA PD000476: P151-R464	MOTIFS HAMMER_PPAM BLIMPS_BLOCKS BLIMPS_PRINTS BLAST_PRODOM
---	-------------	-----	---	---	---

WO 01/90330

PCT/US01/16808

Table 3

4	5332221CD1	477	T57, T116, S123, T119, S124, T120, S121, T122, S123, T124, S125, T126, S127, T128, S129, T130, S131, T132, S133, T134, S135, T136, S137, T138, S139, T140, S141, T142, S143, T144, S145, T146, S147, T148, S149, T150, S151, T152, S153, T154, S155, T156, S157, T158, S159, T160, S161, T162, S163, T164, S165, T166, S167, T168, S169, T170, S171, T172, S173, T174, S175, T176, S177, T178, S179, T180, S181, T182, S183, T184, S185, T186, S187, T188, S189, T190, S191, T192, S193, T194, S195, T196, S197, T198, S199, T200, S201, T202, S203, T204, S205, T206, S207, T208, S209, T210, S211, T212, S213, T214, S215, T216, S217, T218, S219, T220, S221, T222, S223, T224, S225, T226, S227, T228, S229, T230, S231, T232, S233, T234, S235, T236, S237, T238, S239, T240, S241, T242, S243, T244, S245, T246, S247, T248, S249, T250, S251, T252, S253, T254, S255, T256, S257, T258, S259, T260, S261, T262, S263, T264, S265, T266, S267, T268, S269, T270, S271, T272, S273, T274, S275, T276, S277, T278, S279, T280, S281, T282, S283, T284, S285, T286, S287, T288, S289, T290, S291, T292, S293, T294, S295, T296, S297, T298, S299, T300, S301, T302, S303, T304, S305, T306, S307, T308, S309, T310, S311, T312, S313, T314, S315, T316, S317, T318, S319, T320, S321, T322, S323, T324, S325, T326, S327, T328, S329, T330, S331, T332, S333, T334, S335, T336, S337, T338, S339, T340, S341, T342, S343, T344, S345, T346, S347, T348, S349, T350, S351, T352, S353, T354, S355, T356, S357, T358, S359, T360, S361, T362, S363, T364, S365, T366, S367, T368, S369, T370, S371, T372, S373, T374, S375, T376, S377, T378, S379, T380, S381, T382, S383, T384, S385, T386, S387, T388, S389, T390, S391, T392, S393, T394, S395, T396, S397, T398, S399, T400, S401, T402, S403, T404, S405, T406, S407, T408, S409, T410, S411, T412, S413, T414, S415, T416, S417, T418, S419, T420, S421, T422, S423, T424, S425, T426, S427, T428, S429, T430, S431, T432, S433, T434, S435, T436, S437, T438, S439, T440, S441, T442, S443, T444, S445, T446, S447, T448, S449, T450, S451, T452, S453, T454, S455, T456, S457, T458, S459, T460, S461, T462, S463, T464, S465, T466, S467, T468, S469, T470, S471, T472, S473, T474, S475, T476, S477, T478, S479, T480, S481, T482, S483, T484, S485, T486, S487, T488, S489, T490, S491, T492, S493, T494, S495, T496, S497, T498, S499, T500, S501, T502, S503, T504, S505, T506, S507, T508, S509, T510, S511, T512, S513, T514, S515, T516, S517, T518, S519, T520, S521, T522, S523, T524, S525, T526, S527, T528, S529, T530, S531, T532, S533, T534, S535, T536, S537, T538, S539, T540, S541, T542, S543, T544, S545, T546, S547, T548, S549, T550, S551, T552, S553, T554, S555, T556, S557, T558, S559, T560, S561, T562, S563, T564, S565, T566, S567, T568, S569, T570, S571, T572, S573, T574, S575, T576, S577, T578, S579, T580, S581, T582, S583, T584, S585, T586, S587, T588, S589, T590, S591, T592, S593, T594, S595, T596, S597, T598, S599, T600, S601, T602, S603, T604, S605, T606, S607, T608, S609, T610, S611, T612, S613, T614, S615, T616, S617, T618, S619, T620, S621, T622, S623, T624, S625, T626, S627, T628, S629, T630, S631, T632, S633, T634, S635, T636, S637, T638, S639, T640, S641, T642, S643, T644, S645, T646, S647, T648, S649, T650, S651, T652, S653, T654, S655, T656, S657, T658, S659, T660, S661, T662, S663, T664, S665, T666, S667, T668, S669, T670, S671, T672, S673, T674, S675, T676, S677, T678, S679, T680, S681, T682, S683, T684, S685, T686, S687, T688, S689, T690, S691, T692, S693, T694, S695, T696, S697, T698, S699, T700, S701, T702, S703, T704, S705, T706, S707, T708, S709, T710, S711, T712, S713, T714, S715, T716, S717, T718, S719, T720, S721, T722, S723, T724, S725, T726, S727, T728, S729, T730, S731, T732, S733, T734, S735, T736, S737, T738, S739, T740, S741, T742, S743, T744, S745, T746, S747, T748, S749, T750, S751, T752, S753, T754, S755, T756, S757, T758, S759, T760, S761, T762, S763, T764, S765, T766, S767, T768, S769, T770, S771, T772, S773, T774, S775, T776, S777, T778, S779, T780, S781, T782, S783, T784, S785, T786, S787, T788, S789, T790, S791, T792, S793, T794, S795, T796, S797, T798, S799, T800, S801, T802, S803, T804, S805, T806, S807, T808, S809, T810, S811, T812, S813, T814, S815, T816, S817, T818, S819, T820, S821, T822, S823, T824, S825, T826, S827, T828, S829, T830, S831, T832, S833, T834, S835, T836, S837, T838, S839, T840, S841, T842, S843, T844, S845, T846, S847, T848, S849, T850, S851, T852, S853, T854, S855, T856, S857, T858, S859, T860, S861, T862, S863, T864, S865, T866, S867, T868, S869, T870, S871, T872, S873, T874, S875, T876, S877, T878, S879, T880, S881, T882, S883, T884, S885, T886, S887, T888, S889, T890, S891, T892, S893, T894, S895, T896, S897, T898, S899, T900, S901, T902, S903, T904, S905, T906, S907, T908, S909, T910, S911, T912, S913, T914, S915, T916, S917, T918, S919, T920, S921, T922, S923, T924, S925, T926, S927, T928, S929, T930, S931, T932, S933, T934, S935, T936, S937, T938, S939, T940, S941, T942, S943, T944, S945, T946, S947, T948, S949, T950, S951, T952, S953, T954, S955, T956, S957, T958, S959, T960, S961, T962, S963, T964, S965, T966, S967, T968, S969, T970, S971, T972, S973, T974, S975, T976, S977, T978, S979, T980, S981, T982, S983, T984, S985, T986, S987, T988, S989, T990, S991, T992, S993, T994, S995, T996, S997, T998, S999, T1000, S1001, T1002, S1003, T1004, S1005, T1006, S1007, T1008, S1009, T1010, S1011, T1012, S1013, T1014, S1015, T1016, S1017, T1018, S1019, T1020, S1021, T1022, S1023, T1024, S1025, T1026, S1027, T1028, S1029, T1030, S1031, T1032, S1033, T1034, S1035, T1036, S1037, T1038, S1039, T1040, S1041, T1042, S1043, T1044, S1045, T1046, S1047, T1048, S1049, T1050, S1051, T1052, S1053, T1054, S1055, T1056, S1057, T1058, S1059, T1060, S1061, T1062, S1063, T1064, S1065, T1066, S1067, T1068, S1069, T1070, S1071, T1072, S1073, T1074, S1075, T1076, S1077, T1078, S1079, T1080, S1081, T1082, S1083, T1084, S1085, T1086, S1087, T1088, S1089, T1090, S1091, T1092, S1093, T1094, S1095, T1096, S1097, T1098, S1099, T1100, S1101, T1102, S1103, T1104, S1105, T1106, S1107, T1108, S1109, T1110, S1111, T1112, S1113, T1114, S1115, T1116, S1117, T1118, S1119, T1120, S1121, T1122, S1123, T1124, S1125, T1126, S1127, T1128, S1129, T1130, S1131, T1132, S1133, T1134, S1135, T1136, S1137, T1138, S1139, T1140, S1141, T1142, S1143, T1144, S1145, T1146, S1147, T1148, S1149, T1150, S1151, T1152, S1153, T1154, S1155, T1156, S1157, T1158, S1159, T1160, S1161, T1162, S1163, T1164, S1165, T1166, S1167, T1168, S1169, T1170, S1171, T1172, S1173, T1174, S1175, T1176, S1177, T1178, S1179, T1180, S1181, T1182, S1183, T1184, S1185, T1186, S1187, T1188, S1189, T1190, S1191, T1192, S1193, T1194, S1195, T1196, S1197, T1198, S1199, T1200, S1201, T1202, S1203, T1204, S1205, T1206, S1207, T1208, S1209, T1210, S1211, T1212, S1213, T1214, S1215, T1216, S1217, T1218, S1219, T1220, S1221, T1222, S1223, T1224, S1225, T1226, S1227, T1228, S1229, T1230, S1231, T1232, S1233, T1234, S1235, T1236, S1237, T1238, S1239, T1240, S1241, T1242, S1243, T1244, S1245, T1246, S1247, T1248, S1249, T1250, S1251, T1252, S1253, T1254, S1255, T1256, S1257, T1258, S1259, T1260, S1261, T1262, S1263, T1264, S1265, T1266, S1267, T1268, S1269, T1270, S1271, T1272, S1273, T1274, S1275, T1276, S1277, T1278, S1279, T1280, S1281, T1282, S1283, T1284, S1285, T1286, S1287, T1288, S1289, T1290, S1291, T1292, S1293, T1294, S1295, T1296, S1297, T1298, S1299, T1300, S1301, T1302, S1303, T1304, S1305, T1306, S1307, T1308, S1309, T1310, S1311, T1312, S1313, T1314, S1315, T1316, S1317, T1318, S1319, T1320, S1321, T1322, S1323, T1324, S1325, T1326, S1327, T1328, S1329, T1330, S1331, T1332, S1333, T1334, S1335, T1336, S1337, T1338, S1339, T1340, S1341, T1342, S1343, T1344, S1345, T1346, S1347, T1348, S1349, T1350, S1351, T1352, S1353, T1354, S1355, T1356, S1357, T1358, S1359, T1360, S1361, T1362, S1363, T1364, S1365, T1366, S1367, T1368, S1369, T1370, S1371, T1372, S1373, T1374, S1375, T1376, S1377, T1378, S1379, T1380, S1381, T1382, S1383, T1384, S1385, T1386, S1387, T1388, S1389, T1390, S1391, T1392, S1393, T1394, S1395, T1396, S1397, T1398, S1399, T1400, S1401, T1402, S1403, T1404, S1405, T1406, S1407, T1408, S1409, T1410, S1411, T1412, S1413, T1414, S1415, T1416, S1417, T1418, S1419, T1420, S1421, T1422, S1423, T1424, S1425, T1426, S1427, T1428, S1429, T1430, S1431, T1432, S1433, T1434, S1435, T1436, S1437, T1438, S1439, T1440, S1441, T1442, S1443, T1444, S1445, T1446, S1447, T1448, S1449, T1450, S1451, T1452, S1453, T1454, S1455, T1456, S1457, T1458, S1459, T1460, S1461, T1462, S1463, T1464, S1465, T1466, S1467, T1468, S1469, T1470, S1471, T1472, S1473, T1474, S1475, T1476, S1477, T1478, S1479, T1480, S1481, T1482, S1483, T1484, S1485, T1486, S1487, T1488, S1489, T1490, S1491, T1492, S1493, T1494, S1495, T1496, S1497, T1498, S1499, T1500, S1501, T1502, S1503, T1504, S1505, T1506, S1507, T1508, S1509, T1510, S1511, T1512, S1513, T1514, S1515, T1516, S1517, T1518, S1519, T1520, S1521, T1522, S1523, T1524, S1525, T1526, S1527, T1528, S1529, T1530, S1531, T1532, S1533, T1534, S1535, T1536, S1537, T1538, S1539, T1540, S1541, T1542, S1543, T1544, S1545, T1546, S1547, T1548, S1549, T1550, S1551, T1552, S1553, T1554, S1555, T1556, S1557, T1558, S1559, T1560, S1561, T1562, S1563, T1564, S1565, T1566, S1567, T1568, S1569, T1570, S1571, T1572, S1573, T1574, S1575, T1576, S1577, T1578, S1579, T1580, S1581, T1582, S1583, T1584, S1585, T1586, S1587, T1588, S1589, T1590, S1591, T1592, S1593, T1594, S1595, T1596, S1597, T1598, S1599, T1600, S1601, T1602, S1603, T1604, S1605, T1606, S1607, T1608, S1609, T1610, S1611, T1612, S1613, T1614, S1615, T1616, S1617, T1618, S1619, T1620, S1621, T1622, S1623, T1624, S1625, T1626, S1627, T1628, S1629, T1630, S1631, T1632, S1633, T1634, S1635, T1636, S1637, T1638, S1639, T1640, S1641, T1642, S1643, T1644, S1645, T1646, S1647, T1648, S1649, T1650, S1651, T1652, S1653, T1654, S1655, T1656, S1657, T1658, S1659, T1660, S1661, T1662, S1663, T1664, S1665, T1666, S1667, T1668, S1669, T1670, S1671, T1672, S1673, T1674, S1675, T1676, S1677, T1678, S1679, T1680, S1681, T1682, S1683, T1684, S1685, T1686, S1687, T1688, S1689, T1690, S1691, T1692, S1693, T1694, S1695, T1696, S1697, T1698, S1699, T1700, S1701, T1702, S1703, T1704, S1705, T1706, S1707, T1708, S1709, T1710, S1711, T1712, S1713, T1714, S1715, T1716, S1717, T1718, S1719, T1720, S1721, T1722, S1723, T1724, S1725, T1726, S1727, T1728, S1729, T1730, S1731, T1732, S1733, T1734, S1735, T1736, S1737, T1738, S1739, T1740, S1741, T1742, S1743, T1744, S1745, T1746, S1747, T1748, S1749, T1750, S1751, T1752, S1753, T1754, S1755, T1756, S1757, T1758, S1759, T1760, S1761, T1762, S1763, T1764, S1765, T1766, S1767, T1768, S1769, T1770, S1771, T1772, S1773, T1774, S1775, T1776, S1777, T1778, S1779, T1780, S1781, T1782, S1783, T1784, S1785, T1786, S1787, T1788, S1789, T1790, S1791, T1792, S1793, T1794, S1795, T1796, S1797, T1798, S1799, T1800, S1801, T1802, S1803, T1804, S1805, T1806, S1807, T1808, S1809, T1810, S1811, T1812, S1813, T1814, S1815, T1816, S1817, T1818, S1819, T1820, S1821, T1822, S1823, T1824, S1825, T1826, S1827, T1828, S1829, T1830, S1831, T1832, S1833, T1834, S1835, T1836, S1837, T1838, S1839, T1840, S1841, T1842, S1843, T1844, S1845, T1846, S1847, T1848, S1849, T1850, S1851, T1852, S1853, T1854, S1855, T1856, S1857, T1858, S1859, T1860, S1861, T1862, S1863, T1864, S1865, T1866, S1867, T1868, S1869, T1870, S1871, T1872, S1873, T1874, S1875, T1876, S1877, T1878, S1879, T1880, S1881, T1882, S1883, T1884, S1885, T1886, S1887, T1888, S1889, T1890, S1891, T1892, S1893, T1894, S1895, T1896, S1897, T1898, S1899, T1900, S1901, T1902, S1903, T1904, S1905, T1906, S1907, T1908, S1909, T1910, S1911, T1912, S1913, T1914, S1915, T1916, S1917, T1918, S1919, T1920, S1921, T1922, S1923, T1924, S1925, T1926, S1927, T1928, S1929, T1930, S1931, T1932, S1933, T1934, S1935, T1936, S1937, T1938, S1939, T1940, S1941, T1942, S1943, T1944, S1945, T1946, S1947, T1948, S1949, T1950, S1951, T1952, S1953, T1954, S1955, T1956, S1957, T1958, S1959, T1960, S1961, T1962, S1963, T1964, S1965, T1966, S1967, T1968, S1969, T1970, S1971,
---	------------	-----	--

WO 01/90330

PCT/US01/16808

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
5	4574912CB1	1320	1-84 752-1103	2700594F6 (OTABTUT10)	703	1236
				7176695H1 (BRSTTTC01)	87	544
				566451B2	1	419
				1511674T6 (LUNGNOT14)	1209	1891
				1437821T6 (PANANOT08)	1218	1895
6	7475765CB1	2480	1-1453	452757F1 (TLYANOT02)	1289	1920
				1437821F6 (PANANOT08)	483	1099
				70997854V1	1513	2189
				8024941J2	252	800
				70995240V1	1418	2021
7	7475776CB1	2714	1-1807	70996436V1	668	1268
				7739120H1	45	484
				6323951H1 (LUNGDIIN02)	1	151
				70998726V1	2059	2480
				70997215V1	897	1434
8	5332221CB1	1672	1-862	74520931V1	701	1237
				5534850F6 (KIDNNOT25)	477	1145
				71023835V1	1893	2531
				6550760H1 (BRANON02)	1917	2633
				5974444H1 (BRANOT01)	1154	1839
				1680426F6 (STONET01)	2195	2714
				71426048V1	1228	1925
				78903556 (PROSTUT03)	700	1259
				64692956 (BRSTTUT02)	1269	1672
				1712009X14C1 (PROSNOT16)	1	618
				2841527H1 (DRGANOT01)	690	955
				3222761R6 (COLANON03)	976	1484
				1712009X16C1 (PROSNOT16)	288	951

WO 01/90330

PCT/US01/16808

Table 5

Polynucleotide Seq. ID No.	Incyte Project ID	Representative Library
5	4574912CB1	LINGTUT03
6	7475776CB1	KERANOT01
7	7475776CB1	LNOONOT03
8	5332221CB1	COLANOT11

WO 01/90330

PCT/US01/16808

Table 6

Library	Vector	Library Description
LUNG2UT03	PSPORT1	Library was constructed using RNA isolated from lung tumor tissue removed from the left lower lobe of a 69-year-old Caucasian male. The patient had a history of extensive squamous cell carcinoma. Pathology indicated residual tumor, extensive myocardial infarction, prostatic carcinoma. Patient history included severe hyperlipidemia, neoplasm, and tobacco use.
KERANOT01	PBLUESCRIPT	Library was constructed using RNA isolated from neonatal keratinocytes obtained from the leg skin of a spontaneously aborted black male.
LNODNOT03	PINCY	Library was constructed using RNA isolated from lymph node tissue obtained from a 67-year-old Caucasian male during a segmental lung resection and bronchoscopy. On microscopic exam, this tissue was found to be extensively necrotic with 10% viable tumor. Pathology for the associated tumor tissue indicated invasive grade 3-4 squamous cell carcinoma. Patient history included hemangioma. Family history included atherosclerotic coronary artery disease, benign hypertension, congestive heart failure, atherosclerotic coronary artery disease.
COLONNOT11	PSPORT1	Library was constructed using RNA isolated from colon tissue removed from a 60-year-old Caucasian male during a hemicolectomy.

WO 01/90330

PCT/US01/16808

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, tblastn, and tblastx.	Altschul, S.F. et al. (1990) <i>J. Mol. Biol.</i> 215:403-410; Altschul, S.F. et al. (1997) <i>Nucleic Acids Res.</i> 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less <i>Full Length sequences</i> : 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, tfasta, fblast, fblastx, and search.	Pearson, W.R. and D.J. Lipman (1988) <i>Proc. Natl. Acad. Sci. USA</i> 85:2444-2448; Pearson, W.R. (1990) <i>Methods Enzymol.</i> 183:63-98; and Smith, T.F. and M.S. Waterman (1981) <i>Adv. Appl. Math.</i> 2:482-489.	ESTs: fasta E value=1.06E-6 <i>Assembled ESTs</i> : fasta Identity= 95% or greater and Match length=200 bases or greater; fasta E value=1.0E-8 or less <i>Full Length sequences</i> : 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) <i>Nucleic Acids Res.</i> 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) <i>Methods Enzymol.</i> 266:88-105; and Atwood, T.K. et al. (1997) <i>J. Chem. Inf. Comput. Sci.</i> 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) <i>J. Mol. Biol.</i> 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) <i>Nucleic Acids Res.</i> 26:320-322; Durbin, R. et al. (1998) <i>Our World View</i> , in a Nussell, Cambridge Univ. Press, pp. 1-350.	PFAM <i>hsc</i> : Probability value= 1.0E-3 or less <i>Signal peptide hsc</i> : Score= 0 or greater

WO 01/90330

PCT/US01/16808

Table 7 (cont.)

Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite. Gribkov, M. et al. (1988) CABIOS 4:61-66; Gribkov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality scores>CCC-specified "HIGP" 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.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides. Gordon, D. et al. (1998) Genome Res. 8:195-202.	
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation. Nelson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMHMMER	A program that uses a hidden Markov model (HMM) 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. 5:363-371. Somnhammer, E.L. et al. (1998) Proc. Sixth Int. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al. eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, 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 01/90330

PCT/US01/16808

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
5 SEQ ID NO:1-4,
 - b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical
to an amino acid sequence selected from the group consisting of SEQ ID NO:1-4,
 - c) a biologically active fragment of a polypeptide having an amino acid sequence selected
from the group consisting of SEQ ID NO:1-4, and
 - 10 d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from
the group consisting of SEQ ID NO:1-4.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-4.
- 15 3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:5-
20 8.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a
polynucleotide of claim 3.
- 25 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 for producing a polypeptide of claim 1, the method comprising:
 - 30 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.

WO 01/90330

PCT/US01/16808

10. An isolated antibody which specifically binds to a polypeptide of claim 1.
11. An isolated polynucleotide selected from the group consisting of:
- 5 a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:5-8,
- b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:5-8,
- c) a polynucleotide complementary to a polynucleotide of a),
- d) a polynucleotide complementary to a polynucleotide of b), and
- 10 e) an RNA equivalent of a)-d).
12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.
- 15 13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, 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 which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex
- 20 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.
14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.
- 25 15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- 30 b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.
16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable

WO 01/90330

PCT/US01/16808

excipient.

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-4.

5

18. A method for treating a disease or condition associated with decreased expression of functional ATRS, comprising administering to a patient in need of such treatment the composition of claim 16.

10

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

15

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

20

21. A method for treating a disease or condition associated with decreased expression of functional ATRS, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

25

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

30

24. A method for treating a disease or condition associated with overexpression of functional ATRS, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim

WO 01/90330

PCT/US01/16808

1, said method comprising the steps of:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- 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.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said 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.

27. A method for 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.

28. 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) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 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 11 or fragment thereof;

WO 01/90330

PCT/US01/16808

c) quantifying the 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.

5

29. A diagnostic test for a condition or disease associated with the expression of ATRS in a biological sample comprising the steps of:

a) combining the biological sample with an antibody of claim 10, 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.

30. The antibody of claim 10, 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 31. A composition comprising an antibody of claim 10 and an acceptable excipient.

32. A method of diagnosing a condition or disease associated with the expression of ATRS in a subject, comprising administering to said subject an effective amount of the composition of claim 31.

25 33. A composition of claim 31, wherein the antibody is labeled.

34. A method of diagnosing a condition or disease associated with the expression of ATRS in a subject, comprising administering to said subject an effective amount of the composition of claim 33.

30 35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, or an immunogenic fragment thereof, under conditions to elicit an

WO 01/90330

PCT/US01/16808

antibody response;

- b) isolating antibodies from said animal; and
 - c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4.
- 5
26. An antibody produced by a method of claim 35.
37. A composition comprising the antibody of claim 36 and a suitable carrier.
- 10
38. A method of making a monoclonal antibody with the specificity of the antibody of claim 10 comprising:
- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4, 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;
 - d) culturing the hybridoma cells; and
 - e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4.
- 15
39. A monoclonal antibody produced by a method of claim 38.
- 25
40. A composition comprising the antibody of claim 39 and a suitable carrier.
41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.
- 30
42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.
43. A method for detecting a polypeptide having an amino acid sequence selected from the

WO 01/90330

PCT/US01/16808

group consisting of SEQ ID NO:1-4 in a sample, comprising the steps of:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

5 b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4 in the sample.

44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4 from a sample, the method comprising:

10 a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4.

15 45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

20 48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

49. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:5.

25 50. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:6.

51. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:7.

52. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:8.

WO 01/90330

PCT/US01/16808

<110> INCYTE GENOMICS, INC.

YUE, Henry
 TANG, Y. Tom
 PATTERSON, Chandra
 GANDHI, Ameena R.
 TRIBOULEY, Catherine M.
 LEE, Ernestine A.
 YAO, Monique G.
 EANDMAN, Olga
 LU, Dyung Aina M.

<120> AMINOACYL tRNA SYNTHETASES

<130> PI-0109 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/207,248; 60/208,791; 60/210,585

<151> 2000-05-25; 2000-06-01; 2000-06-08

<160> 8

<170> FERL Program

<210> 1

<211> 564

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 4574912CD1

<400> 1

```

Met Leu Arg Thr Thr Arg Gly Pro Gly Leu Gly Pro Pro Leu Leu
1      5      10      15
Gln Ala Ala Leu Gly Leu Gly Arg Ala Gly Trp His Trp Pro Ala
20     25     30
Gly Arg Ala Ala Ser Gly Gly Arg Gly Arg Ala Trp Leu Gln Pro
35     40     45
Thr Gly Arg Glu Thr Gly Val Gln Val Tyr Asn Ser Leu Thr Gly
50     55     60
Arg Lys Glu Pro Leu Ile Val Ala His Ala Glu Ala Ala Ser Trp
65     70     75
Tyr Ser Cys Gly Pro Thr Val Tyr Asp His Ala His Leu Gly His
80     85     90
Ala Cys Ser Tyr Val Arg Phe Asp Ile Ile Arg Arg Ile Leu Thr
95     100    105
Lys Val Phe Gly Cys Ser Ile Val Met Val Met Gly Ile Thr Asp
110    115    120
Val Asp Asp Lys Ile Ile Lys Arg Ala Asn Glu Met Asn Ile Ser
125    130    135
Pro Ala Ser Leu Ala Ser Leu Tyr Glu Glu Asp Phe Lys Gln Asp
140    145    150
Met Ala Ala Leu Lys Val Leu Pro Pro Thr Val Tyr Leu Arg Val
155    160    165
Thr Glu Asn Ile Pro Gln Ile Ile Ser Phe Ile Glu Gly Ile Ile
170    175    180
Ala Arg Gly Asn Ala Tyr Ser Thr Ala Lys Gly Asn Val Tyr Phe
185    190    195
Asp Leu Lys Ser Arg Gly Asp Lys Tyr Gly Lys Leu Val Gly Val
200    205    210

```

WO 01/90330

PCT/US01/16808

```

Val Pro Gly Pro Val Gly Glu Pro Ala Asp Ser Asp Lys Arg His
215 220
Ala Ser Asp Phe Ala Leu Trp Lys Ala Ala Lys Pro Gln Glu Val
230 235
Phe Trp Ala Ser Pro Trp Gly Pro Gly Arg Pro Gly Trp His Ile
245 250
Glu Cys Ser Ala Ile Ala Ser Met Val Phe Gly Ser Gln Leu Asp
260 265
Ile His Ser Gly Gly Ile Asp Leu Ala Phe Pro His His Glu Asn
275 280
Glu Ile Ala Gln Cys Glu Val Phe His Gln Cys Glu Gln Trp Gly
290 295
Asn Tyr Phe Leu His Ser Gly His Leu His Ala Lys Gly Lys Glu
305 310
Glu Lys Met Ser Lys Ser Leu Lys Asn Tyr Ile Thr Ile Lys Asp
320 325
Phe Leu Lys Thr Phe Ser Pro Asp Val Phe Arg Phe Phe Cys Leu
335 340
Arg Ser Ser Tyr Arg Ser Ala Ile Asp Tyr Ser Asp Ser Ala Met
350 355
Leu Gln Ala Gln Gln Leu Leu Leu Gly Leu Gly Ser Phe Leu Glu
365 370
Asp Ala Arg Ala Tyr Met Lys Gly Gln Leu Ala Cys Gly Ser Val
380 385
Arg Glu Ala Met Leu Trp Glu Arg Leu Ser Ser Thr Lys Arg Ala
395 400
Val Lys Ala Ala Leu Ala Asp Asp Phe Asp Thr Pro Arg Val Val
410 415
Asp Ala Ile Leu Gly Leu Ala His His Gly Asn Gly Gln Leu Arg
425 430
Ala Ser Leu Lys Glu Pro Glu Gly Pro Arg Ser Pro Ala Val Phe
440 445
Gly Ala Ile Ile Ser Tyr Phe Glu Gln Phe Phe Glu Thr Val Gly
455 460
Ile Ser Leu Ala Asn Gln Gln Tyr Val Ser Gly Asp Gly Ser Glu
470 475
Ala Thr Leu His Gly Val Val Asp Glu Leu Val Arg Phe Arg Gln
485 490
Lys Val Arg Gln Phe Ala Leu Ala Met Pro Glu Ala Thr Gly Asp
500 505
Ala Arg Arg Gln Gln Leu Leu Glu Arg Gln Pro Leu Leu Glu Ala
515 520
Cys Asp Thr Leu Arg Arg Gly Leu Thr Ala His Gly Ile Asn Ile
530 535
Lys Asp Arg Ser Ser Thr Thr Ser Thr Trp Glu Leu Leu Asp Gln
545 550
Arg Thr Lys Asp Gln Lys Ser Ala Gly
560

```

<210> 2

<211> 477

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7475765CD1

<400> 2

```

Met Leu Gly Val Arg Cys Leu Leu Arg Ser Val Arg Phe Cys Ser
1 5 10 15
Ser Ala Pro Phe Pro Lys His Lys Pro Ser Ala Lys Leu Ser Val

```

WO 01/90330

PCT/US01/16808

Arg Asp Ala Leu	20	25	30
Gly Ala Gln Asn Ala Ser	35	40	45
Ile Gln Gly Trp	50	55	60
Val Arg Ser Val Arg Ser	65	70	75
Gln Lys Glu Val Leu	80	85	90
Phe Leu His Val Asn Asp Gly Ser Ser	95	100	105
Leu Glu Ser Leu Gln Val	110	115	120
Val Ala Asp Ser Gly Leu Asp Ser Arg	125	130	135
Glu Leu Thr Phe Gly Ser	140	145	150
Ser Val Glu Val Gln Gly Gln Leu Ile	155	160	165
Lys Ser Pro Ser Lys Arg	170	175	180
Gln Asn Val Glu Leu Lys Ala Glu Lys	185	190	195
Ile Lys Val Ile Gly Asn	200	205	210
Cys Asp Ala Lys Asp Phe Pro Ile Lys	215	220	225
Tyr Lys Glu Arg His Pro	230	235	240
Leu Glu Tyr Leu Arg Gln Tyr Pro His	245	250	255
Phe Arg Cys Arg Thr Asn	260	265	270
Val Leu Gly Ser Ile Leu Arg Ile Arg	275	280	285
Ser Glu Ala Thr Ala Ala	290	295	300
Ile His Ser Phe Phe Lys Asp Ser Gly	305	310	315
Phe Val His Ile His Thr	320	325	330
Pro Ile Ile Thr Ser Asn Asp Ser Glu	335	340	345
Gly Ala Gly Glu Leu Phe	350	355	360
Gln Leu Glu Pro Ser Gly Lys Leu Lys	365	370	375
Val Pro Glu Glu Asn Phe	380	385	390
Phe Asn Val Pro Ala Phe Leu Thr Val	395	400	405
Ser Gly Gln Leu His Leu	410	415	420
Glu Val Met Ser Gly Ala Phe Thr Gln	425	430	435
Val Phe Thr Phe Gly Pro	440	445	450
Thr Phe Arg Ala Glu Asn Ser Gln Ser	455	460	465
Arg Arg His Leu Ala Glu	470		
Phe Tyr Met Ile Glu Ala Glu Ile Ser			
Phe Val Asp Ser Leu Gln			
Asp Leu Met Gln Val Ile Glu Glu Leu			
Phe Lys Ala Thr Thr Met			
Met Val Leu Ser Lys Cys Pro Glu Asp			
Val Glu Leu Cys His Lys			
Phe Ile Ala Pro Gly Gln Lys Asp Arg			
Leu Glu His Met Leu Lys			
Asn Asn Phe Leu Ile Ile Ser Tyr Thr			
Glu Ala Val Glu Ile Leu			
Lys Gln Ala Ser Gln Asn Phe Thr Phe			
Thr Pro Glu Trp Gly Ala			
Asp Leu Arg Thr Glu His Glu Lys Tyr			
Leu Val Lys His Cys Gly			
Asn Ile Pro Val Phe Val Ile Asn Tyr			
Pro Leu Thr Leu Lys Pro			
Phe Tyr Met Arg Asp Asn Glu Asp Gly			
Pro Gln His Thr Val Ala			
Ala Val Asp Leu Leu Val Pro Gly			
Val Gly Glu Leu Phe Gly Gly			
Gly Leu Arg Glu Glu Arg Tyr His			
Phe Leu Glu Glu Arg Leu Ala			
Arg Ser Gly Leu Thr Glu Val Tyr Gln			
Trp Tyr Leu Asp Leu Arg			
Arg Phe Gly Ser Val Pro His Gly Gly			
Phe Gly Met Gly Phe Glu			
Arg Tyr Leu Gln Cys Ile Leu Gly Val			
Asp Asn Ile Lys Asp Val			
Ile Pro Phe Pro Arg Phe Pro His Ser			
Cys Leu Leu			

WO 01/90330

PCT/US01/16808

<210> 3
 <211> 621
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7475776CD1

<400> 3
 Met Gly Ser Cys Ala Pro Gly Arg Ser His Arg Ala Gly Arg Ala
 1 5 10 15
 Val Asp Ser Pro Cys Leu Ser Phe Ser Arg Ser Gly Asp Val Ile
 20 25 30
 Glu Tyr Leu Leu Lys Asn Gln Trp Phe Val Arg Cys Gln Glu Met
 35 40 45
 Gly Ala Arg Ala Ala Lys Ala Val Glu Ser Gly Ala Leu Glu Leu
 50 55 60
 Ser Pro Ser Phe His Gln Lys Asn Trp Gln His Trp Phe Ser His
 65 70 75
 Ile Gly Asp Trp Cys Val Ser Arg Gln Leu Trp Trp Gly His Gln
 80 85 90
 Ile Pro Ala Tyr Leu Val Val Glu Asp His Ala Gln Gly Glu Glu
 95 100 105
 Asp Cys Trp Val Val Gly Arg Ser Glu Ala Glu Ala Arg Glu Val
 110 115 120
 Ala Ala Glu Leu Thr Gly Arg Pro Gly Ala Glu Leu Thr Leu Glu
 125 130 135
 Arg Asp Pro Asp Val Leu Asp Thr Trp Phe Ser Ser Ala Leu Phe
 140 145 150
 Pro Phe Ser Ala Leu Gly Trp Pro Gln Glu Thr Pro Asp Leu Ala
 155 160 165
 Arg Phe Tyr Pro Leu Ser Leu Leu Glu Thr Gly Ser Asp Leu Leu
 170 175 180
 Leu Phe Trp Val Gly Arg Met Val Met Leu Gly Thr Gln Leu Thr
 185 190 195
 Gly Gln Leu Pro Phe Ser Lys Val Leu His Pro Met Val Arg
 200 205 210
 Asp Arg Gln Gly Arg Lys Met Ser Lys Ser Leu Gly Asn Val Leu
 215 220 225
 Asp Pro Arg Asp Ile Ile Ser Gly Val Glu Met Gln Leu Leu Gln
 230 235 240
 Glu Lys Leu Arg Ser Gly Asn Leu Asp Pro Ala Glu Leu Ala Ile
 245 250 255
 Val Ala Ala Ala Gln Lys Lys Asp Phe Pro His Gly Ile Pro Glu
 260 265 270
 Cys Gly Thr Asp Ala Leu Arg Phe Thr Leu Cys Ser His Gly Val
 275 280 285
 Gln Ala Gly Asp Leu His Leu Ser Val Ser Glu Val Gln Ser Cys
 290 295 300
 Arg His Phe Cys Asn Lys Ile Trp Asn Ala Leu Arg Phe Ile Leu
 305 310 315
 Asn Ala Leu Gly Glu Lys Phe Val Pro Gln Pro Ala Glu Glu Leu
 320 325 330
 Ser Pro Ser Ser Pro Met Asp Ala Trp Ile Leu Ser Arg Leu Ala
 335 340 345
 Leu Ala Ala Gln Glu Cys Glu Arg Gly Phe Leu Thr Arg Glu Leu
 350 355 360
 Ser Leu Val Thr His Ala Leu His His Phe Trp Leu His Asn Leu
 365 370 375
 Cys Asp Val Tyr Leu Glu Ala Val Lys Pro Val Leu Trp His Ser
 380 385 390

WO 01/90330

PCT/US01/16808

```

Pro Arg Pro Leu Gly Pro Pro Gln Val Leu Phe Ser Cys Ala Asp
395 400
Leu Gly Leu Arg Leu Leu Ala Pro Leu Met Pro Phe Leu Ala Glu
410 415
Glu Leu Trp Gln Arg Leu Pro Pro Arg Pro Gly Cys Pro Pro Ala
425 430
Pro Ser Ile Ser Val Ala Pro Tyr Pro Ser Ala Cys Ser Leu Glu
440 445
His Trp Arg Gln Pro Glu Leu Glu Arg Arg Phe Ser Arg Val Gln
455 460
Glu Val Val Gln Val Leu Arg Ala Leu Gln Ala Thr Tyr Gln Leu
470 475
Thr Lys Ala Arg Pro Arg Val Leu Leu Gln Ser Ser Glu Pro Gly
485 490
Asp Gln Gly Leu Phe Glu Ala Phe Leu Glu Pro Leu Gly Thr Leu
500 505
Gly Tyr Cys Gly Ala Val Gly Leu Leu Pro Pro Gly Thr Ala Ala
515 520
Pro Ser Gly Trp Ala Gln Ala Pro Leu Ser Asp Thr Ala Gln Val
530 535
Tyr Met Glu Leu Gln Gly Leu Val Asp Pro Gln Ile Gln Leu Pro
545 550
Leu Leu Ala Ala Arg Arg Tyr Lys Leu Gln Lys Gln Leu Asp Ser
560 565
Leu Thr Ala Arg Thr Pro Ser Glu Gly Glu Ala Gly Thr Gln Arg
575 580
Gln Gln Lys Leu Ser Ser Leu Gln Leu Glu Leu Ser Lys Leu Asp
590 595
Lys Ala Ala Ser His Leu Arg Gln Leu Met Asp Glu Pro Pro Ala
605 610
Pro Gly Ser Pro Glu Leu
620

```

<210> 4

<211> 477

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5332221CD1

<400> 4

```

Met Ala Ala Pro Ile Leu Arg Ser Phe Ser Trp Gly Arg Trp Ser
1 5 10 15
Gly Thr Leu Asn Leu Ser Val Leu Leu Pro Leu Gly Leu Arg Lys
20 25 30
Ala His Ser Gly Ala Gln Gly Leu Leu Ala Ala Gln Lys Ala Arg
35 40 45
Gly Leu Phe Lys Asp Phe Phe Pro Glu Thr Gly Thr Lys Ile Glu
50 55 60
Leu Pro Glu Leu Phe Asp Arg Gly Thr Ala Ser Phe Pro Gln Thr
65 70 75
Ile Tyr Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His Val Gly
80 85 90
His Leu Leu Ala Leu Leu Gly Leu Phe His Leu Gln Arg Ala Gly
95 100 105
His Asn Val Ile Ala Leu Val Gly Gly Ala Thr Ala Arg Leu Gly
110 115 120
Asp Pro Ser Gly Arg Thr Lys Glu Arg Glu Ala Leu Glu Thr Glu
125 130 135
Arg Val Arg Ala Asn Ala Arg Ala Leu Arg Leu Gly Leu Glu Ala

```

WO 01/90330

PCT/US01/16808

```

140      145      150
Leu Ala Ala Asn His Gln Gln Leu Phe Thr Asp Gly Arg Ser Trp
155      160      165
Gly Ser Phe Thr Val Leu Asp Asn Ser Ala Trp Tyr Gln Lys Gln
170      175      180
His Leu Val Asp Phe Leu Ala Ala Val Gly Gly His Phe Arg Met
185      190      195
Gly Thr Leu Leu Ser Arg Gln Ser Val Gln Leu Arg Leu Lys Ser
200      205      210
Pro Glu Gly Met Ser Leu Ala Glu Phe Phe Tyr Gln Val Leu Gln
215      220      225
Ala Tyr Asp Phe Tyr Leu Phe Gln Arg Tyr Gly Cys Arg Val
230      235      240
Gln Leu Gly Gly Ser Asp Gln Leu Gly Asn Ile Met Ser Gly Tyr
245      250      255
Glu Phe Ile Asn Lys Leu Thr Gly Glu Asp Val Phe Gly Ile Thr
260      265      270
Val Pro Leu Ile Thr Ser Thr Thr Gly Ala Lys Leu Gly Lys Ser
275      280      285
Ala Gly Asn Ala Val Trp Leu Asn Arg Asp Lys Thr Ser Pro Phe
290      295      300
Glu Leu Tyr Gln Phe Phe Val Arg Gln Pro Asp Asp Ser Val Glu
305      310      315
Arg Tyr Leu Lys Leu Phe Thr Phe Leu Pro Leu Pro Glu Ile Asp
320      325      330
His Ile Met Gln Leu His Val Lys Glu Pro Glu Arg Arg Gly Pro
335      340      345
Gln Lys Arg Leu Ala Ala Glu Val Thr Lys Leu Val His Gly Arg
350      355      360
Glu Gly Leu Asp Ser Ala Lys Arg Cys Thr Gln Ala Leu Tyr His
365      370      375
Ser Ser Ile Asp Ala Leu Glu Val Met Ser Asp Gln Glu Leu Lys
380      385      390
Glu Leu Phe Lys Glu Ala Pro Phe Ser Glu Phe Phe Leu Asp Pro
395      400      405
Gly Thr Ser Val Leu Asp Thr Cys Arg Lys Ala Asn Ala Ile Pro
410      415      420
Asp Gly Pro Arg Gly Tyr Arg Met Ile Thr Glu Gly Gly Val Ser
425      430      435
Ile Asn His Gln Gln Val Thr Asn Pro Glu Ser Val Leu Ile Val
440      445      450
Gly Gln His Ile Leu Lys Asn Gly Leu Ser Leu Leu Lys Ile Gly
455      460      465
Lys Arg Asn Phe Tyr Ile Ile Lys Trp Leu Gln Leu
470      475

```

<210> 5

<211> 1920

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 4574912CB1

<400> 5

```

tgccgggtgtc actctttatt gcgggggtcca cactgtgggt gctggggccc ctcccactga 60
gggaaggctg agcctctagc cagggctgac atgttgagga ctacgcggg cccaggcctg 120
ggccccccgc tgcctccaggc cgcgctgggc ctggggcggg ctgggtggca ctggcctgcg 180
ggccggggcgg cagcggggggg gcgcggggcgg gcctggctgc agcccaacgg ccgggagacg 240
gggtgtgcagg tgtacaacag cctcaccggg aggaagggaac cctaatcgt ggcgcacgcc 300
gaagccgcct cctgggtatag ctgtggacca actgtatatg atcatgcgca ccttgcccat 360

```

WO 01/90330

PCT/US01/16808

```

gcttgcctcat atgttagatt tgatatactt cgaaggatcc taaccaaggt ttttggatgc 420
agcatagtca tgggtgatggg tattacagat gtagatgata aaatcatcaa aagagccaat 480
gagatgaata ttccccccg ttccctcgcc agtctttatg aggaagactt caagcaggac 540
atggcagccc tgaaggttct cccacccaag gtgtacctga gggtaaccga aaatattctc 600
cagataattt ctttcattga aggaatcatt gctcgtggga acgcttatto aacggcaaaa 660
ggcaatgtct acttcgatct gaagtctaga ggagacaagt atggcaaatl ggtcggcgctg 720
gtccctggtc cagtcggaga gccagcggac tctgacaagt gtcatggcg tgactctgct 780
ctgtggaaag cggccaaacc caaggagggt ttctgggctt ctccctggg acccgggag 840
ccgggtctgg acatcgagtg ctctccatc gctagtatgg tatttggag tcaactggat 900
atccattcag gtgggataga tttagctttt cccatcatg agaacgaat tgcaactgct 960
gaagtctttc atcagtgcca gcagtggga aattatttct tgcattctgg gcaatttgcac 1020
gccaaaggca aagaagaaaa aatgtccaaa tcaataaaga actacattac tattaaggac 1080
tttctgaaga ctttttcccc cgatgtcttc cggttctctt gctcgtggag cagctaccgc 1140
tcagccatcg actacagtga cagcgcctat ctccaagctc agcagctgct cctggggctg 1200
ggctctttcc tggagggagc acgtgcttac atgaaggggc agctggcctg cggctccctc 1260
agggaagcca tgcctgtggg gaggtctctc agcaccaga gggcctgaa ggcggccttg 1320
gcagatgact ttgacacacc cagggtgtgt gatgccatcc tgggcttgc acaccaggg 1380
aatggacagc tcagggcgtc cctgaaggaa cctgaaggcc cgggaagtc tgctgtgttt 1440
ggtgcacata tctcttactt tgaacagttt ttgaaactg ttggaatttc tctggcaaat 1500
caacagtacg ttccaggaga cggcagcgag gctaccttgc atggtgtgtt ggaagagctg 1560
gtgcggtttcc ggacgaagggt ccggcagttt gcctgggcca tgcctggagc caccggggag 1620
gcccggtggc agcgcctctc agaaactcag cccctgtctg aagcatgcca caacctgctg 1680
cggggtctga ctgccacagg catcacatc aaggacagaa gccattacac atccactggg 1740
gaactgctgg atcaaggac aaaagacca aaatcagcgg gctaggagtg gaggcacgct 1800
atgaacctgc tcaagacaag acgcacccat gctctcagg gtcaaggctt tatgttaag 1860
cttctgtctg gggctgctag gtacgacata aagtaaggca accaaccagt aaaaaaaa 1920

```

<210> 6

<211> 2480

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7475765CB1

<400> 6

```

gggctcgagg tgggttcgag tgggttcgag cgggtttctgt tcgtagggtt acaattgtcg 60
cgggtagtgg ggccgccag ctgccaaactg cagctgggttt ctctctggcg ttgctcagt 120
gtcctgtcca ttccgcccg ttgggaagtg ggaagccgaa aaggtcctgt ggggacagag 180
ctgggagctc tgaaggagaa agcgcgcagc cgaagggtctc tggagtgcct tagagcagag 240
ctgcggccgc ggaaggggag cagctgagaa aggaaggccg ctgcaggcgg ggttcgaacc 300
gtgggggtct ggctgtctcc cgggagggcc tggggcggag cgggatgctg ggggtccgct 360
gcctgtctgg gtccgtctgc ttctgttctc cggcccccct ccccaagcac aaaccttcag 420
ccaaactgag cgtgcgggac gctctcgggg ctacagaacgc gactggggag cgcattaaag 480
tccagggatg gattcgttct gtccgatccc agaagggaagt ctltgttctg catgttaagt 540
atgggtctac ttgggaagc ctccaggttg ttgcagattc aggccttgac agtagagaat 600
taacttttgg gacttctgtg gaagtacaag ggcagctgat aaaaagttca cccaaaaggc 660
aaaaatttga actgaaggca gaataaatta aagtatttgg aatttgtgat gccaaagpat 720
tcccatacaa atataaagag aggcattctc tggagtattc gcacaaatat cctcacttta 780
ggtgtaggac taactttctg ggttctatat tgaagattcg cagtgaagcg acagctgcta 840
ttcattcttt cttaaggac agtgcctttg tacatttca taotccaata atcacatcca 900
atgactctga gggagctgga gaacttttct aacttgaaac ttcaggcaaa cttaaggtag 960
ctgaggagaa ttcttcaat gttcctgttt tcttaactgt ctacaggcaa cttoactag 1020
aagtgtatgc aggaagctttt actcaagtgt ttacttttgg tccgacctto cgagctgaaa 1080
attctcagag ccggaggcac ctggcagagt ttatatgat agaagcagag atttcttttg 1140
ttgacagctc tcaagatctt atgcaggtta tagaggaaat gttcaagctc acaaatatga 1200
tggttctctc aaaaatgtct gaagatgttg aactctgtca caaatcata gcaactggcc 1260
aaaaggagac attagaacat atgctaaaaa acaacttttt aatcatttct tatactgaag 1320
cagtggagat cttaaagcaa gcattcccaga acttcaactt taccocagag tgggtgtctg 1380
acctcgggac tgaacatgaa aagtacctgg tgaagcactg tggcaacata cctgtcttct 1440
ttattaatta tcaactaaca ctcaagcctt ttacatgag ggataatgaa gatggccctc 1500

```

WO 01/90330

PCT/US01/16808

```

agcacacggt  tgcgtcgtgt  gatcttctgg  ttctggaggt  tggggaactc  tttggaggag  1560
gctcagagaga  agaacgatac  calttcttag  aggagcgctt  agccagatcg  ggacttacag  1620
aagtctacca  atggtatctg  gaccttcgtc  gatttggatc  tgtgccacat  ggaggttttg  1680
ggatgggatt  tgaacgctac  ctgcagtgc  tcttgggtgt  tgaacaatc  aaagatgtta  1740
tccctttccc  aaggtttcct  cattcatgcc  ttttatagct  ggaagatgg  ttaaggaaaa  1800
gaacccccca  tggcagagac  actgcacatg  attgtgcata  cagcagaatg  catgtttgga  1860
ttttagaat  gacgtttcca  atatgtaatt  gttgtgccat  aagatatcat  agaaaaata  1920
taagtgtgtg  tgattttctt  agaagtttga  gggttattca  cgttaaggtg  agctcccgca  1980
agaagaggtta  cttatagcaa  ggggactctc  aaatccatta  cctcaattaa  gaaatgaaga  2040
aattgaatta  gtctcaaatg  ttcttttaaa  ctctaaaaaa  gaatgagata  atgtatttta  2100
cgtgtcttat  aatcattaaa  tcactccctg  tgtaatttgt  gagaaccatc  tagtagctcg  2160
aaataaata  atgttgcac  ttttccct  tgcacatata  tttgtgata  atctctatct  2220
cattttcagt  acttcattaa  acattgcaga  aaaaaatatt  cctlaaggtc  ttaattgatt  2280
taagaangta  gctattctga  atgaaatct  cctttcattg  aactggata  aaaaatcatg  2340
tttaataact  gttgcttttc  aattttcaaa  gctgttgaga  tattacatta  agtattttca  2400
ctctttaact  actgtgttta  taattgtttt  atatttgatg  ttataattt  gtctaactaa  2460
aatagatttt  ttttaatagac  2480

```

<210> 7

<211> 2714

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7475776CB1

<400> 7

```

cagggcgggtg  caggtgatga  tgatacatct  ggaanaagcaa  aagccaaggt  caggttcagt  60
actcaecatg  gctgtgctcc  ccaaggggca  gtgaagggtga  ctccagctca  cagtcctgcc  120
gatgtctgaga  tgggggcccc  acatggcctg  agccocctga  atgtcattgc  ggaggatggg  180
accatgacct  cctctgccc  ggaactggtg  caggtgggtac  caccctatgt  taccocatcc  240
tttgggggt  cctgtccccc  ctaatccctc  tcttagtttc  ttattctctc  agaggccctc  300
agtctttact  ctgtccgctt  ttctccagg  gtcttcaacg  gtttggagcc  cgggaagaga  360
taagtctgt  gctgagtga  cggggcctat  tccggggcct  ccagaaacca  cccatgggtac  420
tgcacatctg  caggtacctc  attttaactc  ctttactaag  ggtaccacca  aaagggaatg  480
tatggagctt  aaggttgaca  ataggatggg  ctctgcaccc  ctccgttaga  atacgagctc  540
cgtgtcgggt  ttattcgcta  ttgtatccct  agtaccagg  gccctggcatg  gcactggggtc  600
ttgtccccc  gggagaagtc  acagggcccg  aagagcagtg  gactcacct  gtctctcttt  660
cagccgttct  ggggatgtga  tagaatccct  gctgaagaac  cagtggtttg  tccgtcgcca  720
ggaaatgggg  gccgagctg  ccaaggtctg  ggaagtcggg  gccctggagc  tccgtccctc  780
cttcaccag  aagaactggc  agcactgggt  ttcccatatt  ggggactggt  gtgtctcccg  840
gcagctgtgg  tggggccatc  agattccagc  ctacctggtt  gtagaggacc  atgcgcaggg  900
agaagaggac  tgttgggtgg  ttgggcggtc  agaggtctgag  gccagagggt  tagcagcgga  960
actgacaggg  aggcagggg  cagagctgac  cctggagagg  gatcctgatg  tccatagcac  1020
atggttttct  tctgcccgtt  tccccttttc  tgccttgggc  tggccccaag  agaccccaga  1080
cctgtctcgt  ttctacccc  tgtcactttt  ggaacggggc  agcgaccttc  tgcgtttctg  1140
ggtgggcccga  atgttcatgt  tggggaccca  gctcacaggg  cagctgccc  tccagcaggt  1200
gcttctctat  cccatggttc  gggacaggca  gggccgggaag  atgagcaagt  cccctgggaa  1260
tgtgtctggac  ccaagagaca  tcatcagtgg  ggtggagatg  cagtgtctgc  aggaanaagct  1320
gagaagcgga  aatttggacc  ctgcagagct  ggcatttgtg  gctgcagcac  agaaaaggga  1380
ctttctctac  gggatccctg  agtgtgggac  agatgccctg  agattcacac  tctgctccca  1440
tggagttoag  gccggcgact  tgcacctgtc  agtctctgag  gtccagagct  gccgacattt  1500
ctgcaacaag  atctggaatg  ctcttcgctt  tatctcfaat  gctttagggg  agaaatttgt  1560
gccacagcct  gctgaggagc  tgtctccctc  ctcccagatg  gatgcctgga  tccctgagcg  1620
ccttgcctcg  gctgccagg  agtgtgagcg  gggcttctct  acccgagagc  tctcgctcgt  1680
cactcatgcc  ctgcaccact  tctgtcttca  caacctctgt  gacgtctacc  tggaggctgt  1740
gaagcccggt  ctgtggcact  cgccccgcc  cctggggccc  cctcaggtcc  tgttctcctg  1800
cgctgacctc  ggcctccgcc  tctgtggccc  actgatgcoc  ttctgtgctg  aagagctctg  1860
gcagaggctg  ccccccaggc  ctgggttgcc  ccttgccccc  agcatctcgg  ttgccccta  1920
ccctagcgcc  tgcagcttgg  agcactggcg  ccagccagag  ctggagcggt  gcttctcccg  1980
ggtccaagag  gtcgtgcagg  tgcataaggc  tctccaggcc  acgtaccagc  tcaccaagc  2040

```

WO 01/90330

PCT/US01/16808

```

ccggccccga gtgctgtctc agagctcaga gcctggggac cagggcctct tcgaggcctt 2100
cttggagccc ctgggcaccc tgggtctactg tggggctgtg ggctctgttac cccagggcac 2160
agcagctccc tcgggtctggg cccaggtctcc actcagtgac accgctcaag tctacatgga 2220
gctgcagggc ctggtggacc cgcagatcca gctacctctg ttagccgccc gaagggtacaa 2280
gttgacagaag cagcttgata gcttcacagc caggacccca tcagaagggg aggcaggggac 2340
tcaggaggcaa caaaagcttt ctccctcca gctggaaatg tcaaaactgg acaggggagc 2400
ctctcaactc cggcagctga tggatgagcc tccagcccca gggagccgg agctctaaat 2460
catcatcccc atcagttttc ctccctctca gacctgtctt tcaggacaaa cagattttgtc 2520
agctgtcagg gtgcagtggtg acgtcagaga ctatgtgttc catgccttc attgtgttaa 2580
tgaggacaca gactggcttg gtgcagtgga ctgtgtgttc cttgagatgc tccattactc 2640
gcccggcctg cctccacact ggaagtcttg gaatgaggag attgagataa acttttgaaa 2700
tccccaaaaa aaaa
2714

```

<210> 8

<211> 1672

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Inocyte ID No: 5332221CB1

<400> 8

```

gtgggagcct ctgcaatccc agcatteccct tcgtgccgct accaagatgg cggcgcccat 60
cttgoggtcc ttttctctgg gcgggtggtc tggtaacctc aatctctcag tattgttgc 120
cttggggctg cgttaaggccc actcgggcgc tcagggggtta ctggcagcgc agaaggctcg 180
aggtotgttc aaggacttct tcccgagac ggggacgaaa atagagctcc cagagctctt 240
cgaccctggc acggcagatt ttccccaac catttaactg ggcttcgacc ccacggcaga 300
ctcgcttcac gtgggtcatc taacttgctc gctgggcttg ttctatttgc agcagcggg 360
ccacaacgtg atcgcgctgg tgggaggcgc caccggcgcg ctgggagacc cagcggcgg 420
taacaaggaa cgcgagcgcg tgggagacga gcgcgtgcga gccaacgcgc gactctctgc 480
cctagggctt gaggcctctg cggctaatac ccagcagctt ttactgatg ggcgctctg 540
gggcagcttc actgtgctgg acaactcggc ctggtaccag aagcagacc tgggtgactt 600
ctggcggca gtgggggttc acttcgcgat ggggacgctg ctgagccggc agagcgtgca 660
gctggggttc aaggaccccg agggcatgag ctgtggcggg ttcttttacc aggtgtccca 720
ggcctatgac ttctattacc tcttcacgc ttatggatgc aggttcacgc tggcgggac 780
tgatcaacta ggcacacatc tgtccggata tgaagttcac aacaagtiga ctggagaga 840
tgtattttga atcaccgttc ctctaattac aagtacaact ggagcaaac tgggaaagtc 900
tgctggcaac gctgtttggc taacacagga taagacatct ccatttgaat tgtatcaatt 960
ctttgtcagg caaccggacg attcagtgga aaggtacctg aagctgttca ctttctctgc 1020
ccttcacag attgatcata tcatgcagct gcatgtcaaa gagccagaaa ggcgggggtc 1080
tcagaaacga ctggcagcag aagtaacaaa gcttgttcat ggacgagaaa gattggatc 1140
tgctaaaagg tgtacacaag ccttttatca cagtgcata gatgcactgg aggtcatgtc 1200
tgatcaggag ttaaaaagat tgtttaaaga agctccattt tctgaatttt ttctcgatcc 1260
tggacaagat gtccatagata cttgccgcaa agcaaatgcc attccagatg gtccccgagg 1320
gtatcgaatg ataacagaag gcggagtcag cataaatcac caacaagtaa caaatcctga 1380
gagtgthtta attgtggac aacatattct cagaatgga ctttccctac ttaaaataga 1440
aaaaagaat ttctacata taaatggct tcagttgtga tgaaagtcc ttctgtgtgt 1500
ccaaataaac ttacccatca ttcattctca agacctctga aggttggtt ccagaactta 1560
gacctttgtc tatgcacatc agaaaaacag aatgactag gactcagtg gagtaacttc 1620
attattttta tgggcccgtt aataaatatt tgtttaataa aaaaaaaaaa aa 1672

```

【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
29 November 2001 (29.11.2001)

PCT

(10) International Publication Number
WO 01/090330 A3(51) International Patent Classification: C12N 9/00,
A61K 38/53[US/US]: 111 Frederick Court, Mountain View, CA 94043
(US). **BANDMAN, Olga** [US/US]: 366 Anna Avenue,
Mountain View, CA 94043 (US). **LU, Dyung, Alina, M.**
[US/US]: 233 Coy Drive, San Jose, CA 95123 (US).

(21) International Application Number: PCT/US01/16808

(22) International Filing Date: 22 May 2001 (22.05.2001)

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

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/207,248 25 May 2000 (25.05.2000) US
60/208,791 1 June 2000 (01.06.2000) US
60/210,585 8 June 2000 (08.06.2000) 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, 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, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL,
TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.(84) Designated States (*regional*): ARIPO patent (GI, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, 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, HU,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).(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*): **YUE, Henry**
[US/US]: 826 Laik Avenue, Sunnyvale, CA 94087 (US).
TANG, Tom, Y. [US/US]: 4230 Ranwick Court, San
Jose, CA 95118 (US). **PATTERSON, Chandra** [US/US]:
490 Sherwood Way #1, Menlo Park, CA 94025 (US).
GANDHI, Ameena, R. [US/US]: 837 Roble Avenue #1,
Menlo Park, CA 94025 (US). **TRIBOULEY, Catherine,
M.** [FR/US]: 1121 Tennessee Street #5, San Francisco,
CA 94107 (US). **LEE, Ernestine, A.** [US/US]: 624 Kains
Street, Albany, CA 94706 (US). **YAO, Monique, G.**Published:
with international search report(88) Date of publication of the international search report:
8 August 2002For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

WO 01/090330 A3

(54) Title: AMINOACYL tRNA SYNTHETASES

(57) Abstract: The invention provides human aminoacyl tRNA synthetases (ATRS) and polynucleotides which identify and encode ATRS. 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 ATRS.

【国際調査報告】

INTERNATIONAL SEARCH REPORT		National Application No. PCT/US 01/16808
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N9/00 A61K38/53		
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) IPC 7 C12N		
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 practical, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 759 833 A (SHIBA KIYOTAKA ET AL) 2 June 1998 (1998-06-02) the whole document	1-52
A	--- DATABASE EMBL 'Online! EBI Accession No: CAB50310, 6 July 1999 (1999-07-06) HEILIG, R.: "CYSTEINYL-TRNA SYNTHETASE Pyrococcus abyssi" XP002186081 abstract --- -/--	1-7, 11, 12, 45, 49
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents : *A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but 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 such 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 19 December 2001		Date of mailing of the international search report 23.04.02
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentkan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer Schwachtgen, J-L

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/US 01/16808

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	<p>DATABASE EMBL 'Online! EBI Accession No. AK022180, 29 September 2000 (2000-09-29) ISO6AI T., OTSUKI T.: "Homo sapiens cDNA FLJ12118 fis, clone MAMMA1000085, weakly similar to PUTATIVE CYSTEINYL-TRNA SYNTHETASE C29E6.06C (EC 6.1.1.16)" XP002186082 abstract -----</p>	<p>3-5, 11, 12, 49</p>

INTERNATIONAL SEARCH REPORT		International application No. PCT/US 01/16808
Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)		
This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:		
1.	<input checked="" type="checkbox"/> Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely: see FURTHER INFORMATION sheet PCT/ISA/210	
2.	<input type="checkbox"/> Claims 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.	<input type="checkbox"/> Claims 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:		
1.	<input type="checkbox"/> As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.	
2.	<input type="checkbox"/> As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.	
3.	<input type="checkbox"/> 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.	<input checked="" type="checkbox"/> 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.: 45, 49 and 1-44, 46-48, 50-52 (partially)	
Remark on Protest		
<input type="checkbox"/> The additional search fees were accompanied by the applicant's protest.		
<input type="checkbox"/> No protest accompanied the payment of additional search fees.		

International Application No. PCT/US 01/16808

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claims 32 and 34 are directed to a diagnostic method practised on the human body, the search has been carried out and based on the alleged effects of the compound/composition.

Although claims 18, 21, 22 and 24 are directed to a method of treatment of the human body, the search has been carried out and based on the alleged effects of the compound/composition.

International application No.
PCT/US 01/16808

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 45, 49 and 1-44, 46-48, 50-52 (partially)

Isolated polypeptide comprising an amino acid sequence consisting of SEQ ID NO:1.

2. Claims: 46, 50 and 1-45, 47-49, 51, 52 (partially)

Isolated polypeptide comprising an amino acid sequence consisting of SEQ ID NO:2.

3. Claims: 47, 51 and 1-46, 48-50, 52 (partially)

Isolated polypeptide comprising an amino acid sequence consisting of SEQ ID NO:3.

4. Claims: 48, 52 and 1-47, 49-51 (partially)

Isolated polypeptide comprising an amino acid sequence consisting of SEQ ID NO:4.

INTERNATIONAL SEARCH REPORT			
Information on patent family members			International Application No. PCT/US 01/16808
Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5759833 A	02-06-1998	NONE	

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 K 49/00	A 6 1 P 35/00	4 B 0 6 5
A 6 1 P 29/00	A 6 1 P 37/02	4 C 0 8 4
A 6 1 P 35/00	A 6 1 P 43/00	4 C 0 8 5
A 6 1 P 37/02	C 0 7 K 16/40	4 H 0 4 5
A 6 1 P 43/00	C 1 2 N 1/15	
C 0 7 K 16/40	C 1 2 N 1/19	
C 1 2 N 1/15	C 1 2 N 1/21	
C 1 2 N 1/19	C 1 2 N 9/10	
C 1 2 N 1/21	C 1 2 P 21/08	
C 1 2 N 5/10	C 1 2 Q 1/48	Z
C 1 2 N 9/10	C 1 2 Q 1/68	A
C 1 2 P 21/08	G 0 1 N 33/15	Z
C 1 2 Q 1/48	G 0 1 N 33/50	Z
C 1 2 Q 1/68	G 0 1 N 33/53	M
G 0 1 N 33/15	G 0 1 N 33/566	
G 0 1 N 33/50	C 1 2 N 5/00	A
G 0 1 N 33/53	A 6 1 K 37/02	
G 0 1 N 33/566	A 6 1 K 37/48	

(81) 指定国 AP (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, 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, 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, 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, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW

- (72) 発明者 タング、トム・ワイ
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 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 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 0 2 5 ・ メンロパーク ・ # 1 ・ ローブルアベニュー 8 3 7
- (72) 発明者 トリボレー、キャサリーン・エム
アメリカ合衆国カリフォルニア州 9 4 1 0 7 ・ サンフランシスコ ・ # 5 ・ テネシーストリート 1 1 2 1
- (72) 発明者 リー、アーンステーション・エイ
アメリカ合衆国カリフォルニア州 9 4 7 0 6 ・ アルバニー ・ ケインズストリート 6 2 4
- (72) 発明者 ヤオ、モニック・ジー
アメリカ合衆国カリフォルニア州 9 4 0 4 3 ・ マウンテンビュー ・ フレデリックコート 1 1 1
- (72) 発明者 バンドマン、オルガ
アメリカ合衆国カリフォルニア州 9 4 0 4 3 ・ マウンテンビュー ・ アンナアベニュー 3 6 6
- (72) 発明者 リュ、デュング・アリーナ・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

F ターム (参考) 2G045 AA40 DA12 DA13 DA14 FB02
4B024 AA01 AA11 BA10 CA04 CA09 CA11 DA02 DA05 DA06 DA11
DA12 EA02 EA04 GA11 HA12
4B050 CC03 DD01 LL01 LL03
4B063 QA18 QA19 QA20 QQ08 QQ26 QQ43 QR08 QR42 QR56 QS02

	QS25	QS34	QX02							
4B064	AG27	CA10	CA20	CC24	DA01	DA13				
4B065	AA01X	AA57X	AA72X	AA90X	AB01	BA02	CA29	CA44	CA46	
4C084	AA02	AA06	AA07	AA17	BA01	BA02	BA08	BA22	BA23	CA53
	CA56	DC01	NA14	ZB072	ZB112	ZB212	ZB262	ZC022		
4C085	AA02	BB11	EE01	KA03	KA26	LL18				
4H045	AA11	AA20	BA30	CA40	DA76	EA20	EA50	FA72		

专利名称(译)	氨酰基tRNA合成酶		
公开(公告)号	JP2004510407A	公开(公告)日	2004-04-08
申请号	JP2001587126	申请日	2001-05-22
[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	ユエヘンリー タングトムワイ アービズチャンドラエス ガンディーアミーナアール トリボレーキャサリーンエム リーアーンステイーンエイ ヤオモニークジー バンドマンオルガ リュデュングアリーナエム		
发明人	ユエ、ヘンリー タング、トム・ワイ アービズ、チャンドラ・エス ガンディー、アミーナ・アール トリボレー、キャサリーン・エム リー、アーンステイーン・エイ ヤオ、モニーク・ジー バンドマン、オルガ リュ、デュング・アリーナ・エム		
IPC分类号	G01N33/50 A61K38/00 A61K38/43 A61K39/00 A61K45/00 A61K49/00 A61P29/00 A61P35/00 A61P37/02 A61P43/00 C07K16/40 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N9/00 C12N9/10 C12N15/09 C12P21/08 C12Q1/48 C12Q1/68 G01N33/15 G01N33/53 G01N33/566		
CPC分类号	A61P29/00 A61P35/00 A61P37/02 A61P43/00 C12N9/93		
FI分类号	C12N15/00.ZNA.A A61K39/00.Z A61K45/00 A61K49/00.Z A61P29/00 A61P35/00 A61P37/02 A61P43/00.111 C07K16/40 C12N1/15 C12N1/19 C12N1/21 C12N9/10 C12P21/08 C12Q1/48.Z C12Q1/68.A G01N33/15.Z G01N33/50.Z G01N33/53.M G01N33/566 C12N5/00.A A61K37/02 A61K37/48		
F-TERM分类号	2G045/AA40 2G045/DA12 2G045/DA13 2G045/DA14 2G045/FB02 4B024/AA01 4B024/AA11 4B024/BA10 4B024/CA04 4B024/CA09 4B024/CA11 4B024/DA02 4B024/DA05 4B024/DA06 4B024/DA11 4B024/DA12 4B024/EA02 4B024/EA04 4B024/GA11 4B024/HA12 4B050/CC03 4B050/DD01 4B050/LL01 4B050/LL03 4B063/QA18 4B063/QA19 4B063/QA20 4B063/QQ08 4B063/QQ26 4B063/QQ43 4B063/QR08 4B063/QR42 4B063/QR56 4B063/QS02 4B063/QS25 4B063/QS34 4B063/QX02 4B064/AG27 4B064/CA10 4B064/CA20 4B064/CC24 4B064/DA01 4B064/DA13 4B065/AA01X 4B065/AA57X 4B065/AA72X 4B065/AA90X 4B065/AB01 4B065/BA02 4B065/CA29 4B065/CA44 4B065/CA46 4C084/AA02 4C084/AA06 4C084/AA07 4C084/AA17 4C084/BA01 4C084/BA02 4C084/BA08 4C084/BA22 4C084/BA23 4C084/CA53 4C084/CA56 4C084/DC01 4C084/NA14 4C084/ZB072 4C084/ZB112 4C084/ZB212 4C084/ZB262 4C084/ZC022 4C085/AA02 4C085/BB11 4C085/EE01 4C085/KA03 4C085/KA26 4C085/LL18 4H045/AA11 4H045/AA20 4H045/BA30 4H045/CA40 4H045/DA76 4H045/EA20 4H045/EA50 4H045/FA72		
优先权	60/207248 2000-05-25 US 60/208791 2000-06-01 US 60/210585 2000-06-08 US		

摘要(译)

本发明提供了人氨酰基tRNA合成酶 (ATRS) 和鉴定和编码ATRS的多核苷酸。 本发明还提供表达载体，宿主细胞，抗体，激动剂和拮抗剂。 本发明还提供了用于诊断，治疗或预防与ATRS异常表达有关的疾病的方法。