

(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2004-507212

(P2004-507212A)

(43) 公表日 平成16年3月11日(2004.3.11)

(51) Int. Cl. ⁷		F I		テーマコード (参考)	
C 1 2 N	15/09	C 1 2 N	15/00	Z N A A	2 G O 4 5
A O 1 K	67/027	A O 1 K	67/027		4 B O 2 4
A 6 1 K	38/00	A 6 1 K	45/00		4 B O 6 3
A 6 1 K	45/00	A 6 1 P	1/00		4 B O 6 4
A 6 1 P	1/00	A 6 1 P	1/04		4 B O 6 5
		審査請求	未請求	予備審査請求	有 (全 251 頁) 最終頁に続く

(21) 出願番号	特願2001-565896 (P2001-565896)	(71) 出願人	301005050 インサイト・ゲノミックス・インコーポレイテッド アメリカ合衆国カリフォルニア州94304・パロアルト・ポータードライブ 3160
(86) (22) 出願日	平成13年3月1日 (2001.3.1)	(74) 代理人	100089266 弁理士 大島 陽一
(85) 翻訳文提出日	平成14年9月2日 (2002.9.2)	(72) 発明者	ラル、ブリーティ アメリカ合衆国カリフォルニア州95054・サンタクララ・ラスドライブ 2382
(86) 国際出願番号	PCT/US2001/006814		
(87) 国際公開番号	W02001/066742		
(87) 国際公開日	平成13年9月13日 (2001.9.13)		
(31) 優先権主張番号	60/186, 854		
(32) 優先日	平成12年3月3日 (2000.3.3)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/188, 384		
(32) 優先日	平成12年3月10日 (2000.3.10)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/190, 453		
(32) 優先日	平成12年3月17日 (2000.3.17)		
(33) 優先権主張国	米国 (US)		最終頁に続く

(54) 【発明の名称】 Gタンパク質共役受容体

(57) 【要約】

本発明は、ヒトGタンパク質共役受容体 (G C R E C) と、G C R E C を同定及びコードするポリヌクレオチドとを提供する。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提供する。更に、本発明は、G C R E C の異常発現に関連する疾患の診断・治療・予防方法を提供する。

【特許請求の範囲】

【請求項 1】

単離されたポリペプチドであって、

(a) SEQ ID NO: 1 乃至 SEQ ID NO: 21 (SEQ ID NO: 1 - 21) からなる群から選択されたアミノ酸配列と、

(b) SEQ ID NO: 1 - 21 からなる群から選択されたアミノ酸配列と少なくとも 90% の配列同一性を有する天然のアミノ酸配列と、

(c) SEQ ID NO: 1 - 21 からなる群から選択されたアミノ酸配列の生物学的に活性な断片と、

(d) SEQ ID NO: 1 - 21 からなる群から選択されたアミノ酸配列の免疫原性断片とで構成される群から選択されたアミノ酸配列を含むことを特徴とする単離されたポリペプチド。 10

【請求項 2】

SEQ ID NO: 1 - 21 からなる群から選択された請求項 1 の単離されたポリペプチド。

【請求項 3】

請求項 1 のポリペプチドをコードする単離されたポリヌクレオチド。

【請求項 4】

請求項 2 のポリペプチドをコードする単離されたポリヌクレオチド。

【請求項 5】

SEQ ID NO: 22 - 42 からなる群から選択された請求項 4 の単離されたポリヌクレオチド。 20

【請求項 6】

請求項 3 の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。

【請求項 7】

請求項 6 の組換えポリヌクレオチドで形質転換された細胞。

【請求項 8】

請求項 6 の組換えポリヌクレオチドを含む遺伝子組換え生物。

【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ポリペプチドを回収するステップとを含むことを特徴とする請求項 1 のポリペプチドの生産方法。

【請求項 10】

請求項 1 のポリペプチドに特異的に結合する単離された抗体。

【請求項 11】

単離されたポリヌクレオチドであって、 40

(a) SEQ ID NO: 22 - 42 からなる群から選択されたポリヌクレオチド配列と、

(b) SEQ ID NO: 22 - 42 からなる群から選択されたポリヌクレオチド配列と少なくとも 90% の配列同一性を有する天然のポリヌクレオチド配列と、

(c) 前記 (a) に相補的なポリヌクレオチド配列と、

(d) 前記 (b) に相補的なポリヌクレオチド配列と、

(e) 前記 (a) 乃至 (d) の RNA 等価物とで構成される群から選択されたポリヌクレオチド配列を含む単離されたポリヌクレオチド。

【請求項 12】

請求項 11 のポリヌクレオチドの少なくとも 60 個の連続するヌクレオチドを含む単離さ 50

れたポリヌクレオチド。

【請求項 13】

サンプルにおいて、請求項 11 に記載のポリヌクレオチド配列を有する標的ポリヌクレオチドを検出する方法であって、

(a) 前記サンプルをプローブでハイブリダイズするステップであって、前記プローブが、前記サンプル内の前記標的ポリヌクレオチドと相補的な配列を含む少なくとも 20 個の連続するヌクレオチドを含み、前記プローブと前記標的ポリヌクレオチドまたはその断片との間でハイブリダイゼーション複合体が形成される条件下で、前記プローブが前記標的ポリヌクレオチドと特異的にハイブリダイズする、該ステップと、

(b) 前記ハイブリダイゼーション複合体が存在するか否かを検出し、存在する場合には随意選択でその収量を測定するステップとを含むことを特徴とする標的ポリヌクレオチドの検出方法。 10

【請求項 14】

前記プローブが少なくとも 60 個の連続する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】

サンプルにおいて、請求項 11 のポリヌクレオチド配列を有する標的ポリヌクレオチドを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ステップと、

(b) 増幅された前記標的ポリヌクレオチドまたはその断片が存在するか否かを検出し、存在する場合には随意選択でその収量を測定するステップとを含むことを特徴とする標的ポリヌクレオチドの検出方法。 20

【請求項 16】

有効量の請求項 1 のポリペプチド及び医薬的に容認できる賦形剤を含む組成物。

【請求項 17】

前記ポリペプチドが、SEQ ID NO: 1 - 21 からなる群から選択されたアミノ酸配列を含むことを特徴とする請求項 16 の組成物。

【請求項 18】

機能的な GCR E C (新規の G タンパク質共役受容体) 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 のポリペプチドのアゴニストとして効果的な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 のスクリーニング方法によって同定されたアゴニスト化合物及び医薬的に容認できる賦形剤を含む組成物。 40

【請求項 21】

機能的な GCR E C 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 のポリペプチドのアンタゴニストとして効果的な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 50

とするスクリーニング方法。

【請求項 23】

請求項 22 のスクリーニング方法によって同定されたアンタゴニスト化合物及び医薬的に容認できる賦形剤を含む組成物。

【請求項 24】

機能的な G C R E C 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 のポリペプチドに特異的に結合する化合物をスクリーニングする方法であって、 10

(a) 請求項 1 のポリペプチドを好適な条件下で少なくとも 1 つの試験化合物と結合させるステップと、

(b) 請求項 1 のポリペプチドと前記試験化合物との結合を検出して、請求項 1 のポリペプチドと特異的に結合する化合物を同定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 26】

請求項 1 のポリペプチドの活性を変化させ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 のポリペプチドを、その活性が許容される条件下で少なくとも 1 つの試験化合物と結合させるステップと、

(b) 前記試験化合物の存在下での請求項 1 のポリペプチドの活性を評価するステップと 20

(c) 前記試験化合物の存在下での請求項 1 のポリペプチドの活性と、前記試験化合物の不在下での請求項 1 のポリペプチドの活性とを比較するステップとを含み、

前記試験化合物の存在下での請求項 1 のポリペプチドの活性の変化が、請求項 1 のポリペプチドの活性を変化させる化合物の存在を示唆すること特徴とするスクリーニング方法。

【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、 30

(b) 前記標的ポリヌクレオチドの発現の変化を検出するステップと、

(c) 様々な量の前記化合物の存在下での前記標的ポリヌクレオチドの発現と、前記化合物の不在下での前記標的ポリヌクレオチドの発現とを比較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 28】

試験化合物の毒性を評価する方法であって、

(a) 核酸を含む生体サンプルを前記試験化合物で処理するステップと、

(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、 40

(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) 前記処理した生体サンプルにおけるハイブリダイゼーション複合体の収量を、未処理の生体サンプルにおけるハイブリダイゼーション複合体の収量と比較するステップとを含み、

前記処理した生体サンプルにおけるハイブリダイゼーション複合体の収量の差が試験化合物の毒性を示唆することを特徴とする試験化合物の毒性評価方法。

【発明の詳細な説明】

【0001】

(技術分野)

本発明は、Gタンパク質共役受容体の核酸配列、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増殖異常、神経の疾患、心血管疾患、胃腸疾患、自己免疫/炎症性の疾患、代謝障害、およびウイルス感染の診断・治療・予防に関する。本発明はさらに、Gタンパク質共役受容体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影響についての評価に関する。

【0002】

(発明の背景)

シグナル伝達は、細胞外のシグナルに細胞が応答することによる一般的なプロセスである。細胞膜を通過するシグナル伝達は、例えば、ホルモン、神経伝達物質、または成長因子などのシグナル分子が細胞膜受容体に結合することで始まる。このようにして活性化された受容体が、転写因子などの細胞内標的分子の活性化で終了する細胞内の生化学的なカスケードを引き起こす。このシグナル伝達のプロセスは、細胞増殖、分化、および遺伝子転写を含む全てのタイプの細胞機能を調節する。最も大きい遺伝子ファミリーの1つによってコードされるGタンパク質共役受容体は、細胞膜を通過する細胞外シグナルの伝達において重要な役割を演じることが分かった。GPCRは、成功した治療標的であると過去に証明された。

【0003】

GPCRは膜内在性タンパク質であって、1つになって逆平行ヘリックスの束を形成する7つの疎水性の膜貫通ドメインによって特徴付けられる。GPCRの大きさは、アミノ酸400以下から1000以上の範囲である(Strosberg, A. D. (1991) Eur. J. Biochem. 196: 1-10; Coughlin, S. R. (1994) Curr. Opin. Cell Biol. 6: 191-197)。GPCRのアミノ末端はその長さが可変長であって細胞外に存在し、かつグリコシル化している場合が多い。カルボキシル末端は細胞質に存在し、通常はリン酸化している。細胞外ループが細胞内ループと交互し、膜貫通ドメインを繋いでいる。第2の細胞外ループと第3の細胞外ループとを連結しているシステインジスルフィド架橋が、アゴニストやアンタゴニストと相互作用し得る。GPCRの最も保存されたドメインは、膜貫通ドメインおよび初めの2つの細胞質ループである。膜貫通ドメインは、受容体の構造的および機能的特長に部分的に寄与している。大抵の場合、ヘリックスの束がリガンド結合ポケットを形成する。細胞外N末端セグメント、または3つの細胞外ループの内1或いは複数の細胞外ループが、リガンドとの結合に参加し得る。リガンドが結合すると、受容体は、その細胞内部分において構造変化が起こり活性化する。次に、この活性化された受容体の大きな第3の細胞内ループが、更なる細胞内のシグナル伝達作用を仲介する多量体であるグアニンヌクレオチド結合Gタンパク質複合体と相互作用する。この細胞内のシグナル伝達作用には、サイクリックAMP(cAMP)、ホスホリパーゼC、およびイノシトール三リン酸などのセカンドメッセンジャーの活性化、並びに活性化されたGPCRとイオンチャネルタンパク質との相互作用が含まれる(Watson, S. および S. Arkininstall (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 2-6; Bolander, F. F. (1994) Molecular Endocrinology, Academic Press, San Diego CA, pp. 162-176; Baldwin, J. M. (1994) Curr. Opin. Cell Biol. 6: 180-190を参照)。

【0004】

GPCRには、感覚シグナルメディエータ(例えば、光および嗅覚刺激分子)と、アデニン、-アミノ酪酸(GABA)、肝細胞成長因子、メラノコルチン、ニューロペプチドY、オピオイドペプチド、オプシン、ソマトスタチン、タキキニン、血管作用性腸管ポリペプチドファミリー、およびバソプレシンと、生体アミン(例えば、ドーパミン、エピ

10

20

30

40

50

ネフリンおよびノルエピネフリン、ヒスタミン、グルタミン酸（代謝調節効果）、アセチルコリン（ムスカリン効果）、およびセロトニン）と、ケモカインと、炎症の脂質メディエータ（例えば、プロスタグランジンおよびプロスタノイド、血小板活性化因子、およびロイコトリエン）と、ペプチドホルモン（例えば、ボンベシン、ブラジキニン、カルシトニン、C5aアナフィラトキシン、エンドセリン、卵胞刺激ホルモン（FSH）、性腺刺激ホルモン放出ホルモン（GnRH）、ニューロキニン、甲状腺刺激ホルモン放出ホルモン（TRH）、オキシトシン）とが含まれる。分かっていない刺激に対して受容体として作用するGPCRはオーファン受容体と呼ばれる。

【0005】

GPCRファミリーの多様性は、択一的スプライシングによって更に広がる。多くのGPCR遺伝子はイントロンを含み、スプライスバリエーションが見つかったこのような受容体は現在30以上存在する。タンパク質C末端に最も多くの変異が存在する。N末端および細胞質ループにおける変異もしばしば見られるが、細胞外ループ若しくは膜貫通ドメインに於ける変異は一般的でない。受容体の中には、2ヶ所以上で変化が起こるものもある。スプライスバリエーションは、分布、シグナル伝達、結合、調節、およびリガンド結合のプロファイルから判断すると、機能的に異なっていると思われる（Kilpatrick, G. J. ら（1999）*Trends Pharmacol. Sci.* 20 : 294 - 301）。

【0006】

GPCRは、ロドプシン様受容体サブファミリー、セレクトイン様受容体サブファミリー、および代謝調節型グルタミン酸受容体サブファミリーの3つの大きなサブファミリーに分類できる。これらのGPCRサブファミリーのメンバーは、類似の機能および特徴的な7回膜貫通構造を共有するが、アミノ酸配列は多様である。最も大きなファミリーは、ホルモン、神経伝達物質、および光を含む多様な細胞外のシグナルを伝達するロドプシン様GPCRからなる。ロドプシンは、動物網膜に見られる感光性GPCRである。脊椎動物では、ロドプシン分子は、光受容（杆状体）細胞に見られる膜スタック（membranous stack）に埋め込まれている。各ロドプシン分子は光の光子に反応して、細胞膜ナトリウムチャンネルが閉止するレベルにcGMPを低下させる。このようにして、視覚シグナルが神経インパルスに変換される。他のロドプシン様GPCRは、神経伝達物質に対して直接反応する。これらのGPCRには、アドレナリンの受容体（アドレナリン受容体）、アセチルコリンの受容体（ムスカリン様受容体）、アデノシンの受容体、ガラニンの受容体、およびグルタミン酸の受容体（N-メチル-D-アスパラギン酸/NMDA受容体）が含まれる（Watson, S. および S. Arkinstall（1994）*The G-Protein Linked Receptor Facts Book*, Academic Press, San Diego CA, pp. 7 - 9, 19 - 22, 32 - 35, 130 - 131, 214 - 216, 221 - 222 ; Habert-Ortoli, E. ら（1994）*Proc. Natl. Acad. Sci. USA* 91 : 9780 - 9783を参照）。

【0007】

ガラニン受容体は、インスリン、アセチルコリン、セロトニンおよびノルアドレナリンの分泌を阻害し、プロラクチンおよび成長ホルモンの放出を刺激する神経内分泌ペプチドガラニンの活性を仲介する。ガラニン受容体は、摂食障害、苦痛、抑うつ症、およびアルツハイマー病に關与する（Kask, K. ら（1997）*Life Sci.* 60 : 1523 - 1533）。他の神経系ロドプシン様GPCRには、発生および神経病理に役割を持つと考えられるリゾホスファチジン酸および他のリゾリン脂質の受容体の増大するファミリーが含まれる（Chun, J. ら（1999）*Cell Biochem. Biophys.* 30 : 213 - 24）。

【0008】

GPCRの最も大きいファミリーである嗅覚受容体は、ロドプシン様GPCRのファミリーのメンバーでもある。これらの受容体は、匂い物質シグナルを伝達する。様々な匂いを

区別するために多数の異なる嗅覚受容体が必要である。それぞれの嗅覚ニューロンは唯一種類の嗅覚受容体を発現し、異なる受容体を発現するニューロンの異なる空間領域が鼻腔に存在する。例えば、ラット脳ライブラリから単離されたRA1c受容体は、脳の特定の領域および嗅上皮の限定された領域にのみ発現することが分かった(Raming, K. ら (1998) *Receptors Channels* 6 : 141 - 151)。しかしながら、嗅覚様受容体の発現は嗅覚組織に限定されてはいない。例えば、典型的なGPCRの特徴を有する嗅覚様受容体をコードする3つのラット遺伝子は、味覚および嗅覚組織においてのみならず、男性生殖組織にも発現パターンが見られた(Thomas, M. B. ら (1996) *Gene* 178 : 1 - 5)。

【0009】

セレクチン様GPCRサブファミリーのメンバーは、リガンドとして、セレクチン、カルシトニン、グルカゴン、成長ホルモン放出ホルモン、副甲状腺ホルモン、および血管作用性小腸ペプチドなどのペプチドホルモンを有する。例えば、セレクチン受容体は、膵臓および小腸における酵素およびイオンの分泌を刺激するペプチドホルモンであるセレクチンに応答する(Watson, 前出, pp. 278 - 283)。セレクチン受容体は、アミノ酸約450の長さであり、胃腸細胞の細胞膜に見られる。セレクチンがその受容体と結合すると、cAMPの産生が刺激される。

【0010】

炎症反応および免疫反応に關与するセレクチン様GPCRの例には、EGFモジュールを含むムチン様ホルモン受容体(Emr1)およびCD97受容体タンパク質がある。これらのGPCRは、最近特徴付けられたEGF-TM7受容体サブファミリーのメンバーである。これらの7回膜貫通型ホルモン受容体は、*in vivo*でヘテロ二量体として存在し、3個から7個の潜在的なカルシウム結合EGF様モチーフを含む。CD97は、主に白血球で発現され、活性化されたB細胞上およびT細胞上で著しくアップレギュレートされる(McKnight, A. J. および S. Gordon (1998) *J. Leukoc. Biol.* 63 : 271 - 280)。

【0011】

第3のGPCRサブファミリーは、代謝調節型グルタミン酸受容体ファミリーである。グルタミン酸は、中枢神経系における主な興奮性神経伝達物質である。代謝調節型グルタミン酸受容体は、細胞内のエフェクターの活性を調節し、長期に亘る増強作用に關与する(Watson, 前出, p. 130)。カルシウムイオンの細胞外濃度の変化を検出するCa²⁺感受性受容体は、カルシウム結合に關与し得る酸性アミノ酸のクラスターを含む大きな細胞外のドメインを有する。代謝調節型グルタミン酸受容体ファミリーはまた、フェロモン受容体、GABA_B受容体、および味覚受容体を含む。

【0012】

他のGPCRのファミリーには、哺乳動物嗅覚受容体遺伝子に僅かに關連する線虫(*Caenorhabditis elegans* および *Caenorhabditis briggsae*)に見られる2つの化学受容体遺伝子群を含む。細胞膜上の結合因子(mating factor)に対する応答に關与する酵母フェロモン受容体であるSTE2およびSTE3は、粘菌(*Dictyostelium discoideum*)に由来するcAMP受容体と同様に、それら自体の7回膜貫通シグネチャ(signature)を有する、この7回膜貫通シグネチャは、個々の細胞の凝集を調節し、発生的に制御された様々な遺伝子の発現を調節すると考えられている。

【0013】

機能の低下や恒常的な活性を引き起こし得るGPCR変異体は、様々なヒトの疾患に關連する(Coughlin, 前出)。例えば、網膜色素変性症は、ロドプシン遺伝子の変異から生じ得る。更に、甲状腺刺激ホルモン受容体における体細胞活性化突然変異は、機能亢進性甲状腺腫を引き起こすと報告され、恒常的な活性を受けやすいある種のGPCRがプロトオンコジーンとして振る舞い得ることを示唆するものである(Parma, J. ら (1993) *Nature* 365 : 649 - 651)。黄体形成ホルモン(

10

20

30

40

50

思春期早発症)、バソプレシンV2(X染色体性の腎性糖尿病)、グルカゴン(糖尿病および高血圧)、カルシウム(上皮小体機能亢進症、低カルシウム尿症、高カルシウム血症)、副甲状腺ホルモン(短肢矮小発育症)、 β_3 -アドレナリン受容体(肥満症、非インスリン依存性糖尿病)、成長ホルモン放出ホルモン(小人症)、および副腎皮質刺激ホルモン(糖質コルチコイド欠乏症)のリガンドのGPCR受容体も、ヒトの疾患に関連する変異体を含む(Wilson, S. ら(1998) Br. J. Pharmacol. 125: 1387-1392; Stadel, J. M. ら(1997) Trends Pharmacol. Sci. 18: 430-437)。GPCRもまた、抑うつ症、分裂病、不眠症、高血圧症、不安症、ストレス、腎不全、および幾つかの心血管疾患に關与する(Horn, F. および G. Vriend(1998) J. Mol. Med. 76: 464-468)。

10

【0014】

加えて、GPCRを活性化する或いは抑制する数百種の新規の薬剤がこの20年間に見出された。これらの薬剤の治療標的は、癌、骨粗鬆症、および子宮内膜症はもちろん心血管、胃腸、および中枢神経系の疾患を含む疾患および障害の広い範囲に及ぶ(Wilson, 前出; Stadel, 前出)。例えば、パーキンソン病の治療にドーパミンアゴニストL-ドーパが用いられ、分裂病および初期のハンチントン病の治療にドーパミンアンタゴニストが用いられる。喘息、高血圧、その他の心血管疾患、および不安症の治療にアドレナリン受容体のアゴニストおよびアンタゴニストが用いられ、緑内障および頻脈の治療にムスカリン様アゴニストが用いられ、偏頭痛の治療にセロトニン5HT_{1D}アンタゴニストが用いられ、アレルギーおよびアナフィラキシー反応、枯草熱、痒み、および動揺病にヒスタミンH₁アンタゴニストが用いられる。

20

【0015】

近年の研究から、糖尿病、肥満症、および骨粗鬆症を含む代謝障害の治療において、GPCRが将来の治療薬としての可能性があることが分かった。例えば、腎性糖尿病を引き起こすV2バソプレシン受容体変異体は、突然変異を含む領域に渡るC末端V2受容体ペプチドの同時発現により、in vitroでその機能が助けられ得る。このことは、疾患治療の新しい方法の可能性を示している(Schoneberg, T. ら(1996) EMBO J. 15: 1283-1291)。メラノコルチン-4-受容体(MC4R)における突然変異は、ヒトの体重調節および肥満に關与する。これらのMC4R変異体は、バソプレシンV2受容体変異体と同様に、細胞膜への輸送における欠陥(Ho, G. および R. G. MacKenzie(1999) J. Biol. Chem. 274: 35816-35822)であるため、同様の方法で治療できるであろう。副甲状腺ホルモン(PTH)の1型受容体は、血流におけるカルシウム恒常性のPTH依存性の調節を仲介するGPCRである。PTHと受容体との相互作用の研究により、骨粗鬆症の治療のための新規のPTH受容体リガンドの開発が可能となるであろう(Mannstadt, M. ら(1999) Am. J. Physiol. 277: F665-F675)。

30

【0016】

GPCRのケモカイン受容体群は、炎症性および感染性の疾患に用いられる潜在的な治療薬である(Locati, M. および P. M. Murphy(1999) Annu. Rev. Med. 50: 425-440を参照)。ケモカインは、白血球輸送、造血、および脈管形成の調節における細胞内シグナルとして作用する小さなポリペプチドである。マウスにおいて様々な標的ケモカインを破壊する実験によって、これらの受容体が病的な炎症や多発性硬化症などの自己免疫疾患において役割を果たしていることが示された。ケモカイン受容体はまた、感染を促進するためにヘルペスウイルスおよびヒト免疫不全ウイルス(HIV-1)を含む感染因子に利用される。HIV-1によるT細胞の感染のための補助受容体として作用するケモカイン受容体CCR5の切断型がAIDSに対する抵抗性を持つようになることから、CCR5アンタゴニストがAIDSの進行の予防に有用であると考えられる。

40

50

【0017】

新規のGタンパク質共役受容体、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、増殖異常、神経の疾患、心血管疾患、胃腸疾患、自己免疫/炎症性の疾患、代謝障害、およびウイルス感染の診断・治療・予防において有用であり、また、Gタンパク質共役受容体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影響についての評価にも有用である。

【0018】

(発明の要約)

本発明は、総称して「GCREC」、個別にはそれぞれ「GCREC-1」、「GCREC-2」、「GCREC-3」、「GCREC-4」、「GCREC-5」、「GCREC-6」、「GCREC-7」、「GCREC-8」、「GCREC-9」、「GCREC-10」、「GCREC-11」、「GCREC-12」、「GCREC-13」、「GCREC-14」、「GCREC-15」、「GCREC-16」、「GCREC-17」、「GCREC-18」、「GCREC-19」、「GCREC-20」、および「GCREC-21」と呼ぶGタンパク質共役受容体である精製されたポリペプチドを提供する。本発明の一実施態様では、(a)SEQ ID NO: 1-21からなる一群から選択されたアミノ酸配列と、(b)SEQ ID NO: 1-21からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c)SEQ ID NO: 1-21からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d)SEQ ID NO: 1-21とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含む単離されたポリペプチドを提供する。別法では、SEQ ID NO: 1-21のアミノ酸配列を含む単離されたポリペプチドを提供する。

【0019】

更に本発明は、(a)SEQ ID NO: 1-21からなる一群から選択されたアミノ酸配列と、(b)SEQ ID NO: 1-21からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c)SEQ ID NO: 1-21からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d)SEQ ID NO: 1-21からなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドをコードする単離されたポリヌクレオチドを提供する。別法では、このポリヌクレオチドは、SEQ ID NO: 1-21からなる一群から選択されたポリペプチドをコードする。別法では、このポリヌクレオチドは、SEQ ID NO: 22-42からなる一群から選択される。

【0020】

更に、本発明は、(a)SEQ ID NO: 1-21からなる一群から選択されたアミノ酸配列と、(b)SEQ ID NO: 1-21からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c)SEQ ID NO: 1-21からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d)SEQ ID NO: 1-21からなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含む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を提供する。別法では、本発明は、この組換えポリヌクレオチドで形質転換された細胞を提供する。更なる別法では、本発明は、この組換えポリヌクレオチドを含む遺伝子組換え生物を提供する。

【0021】

また、本発明は、(a)SEQ ID NO: 1-21からなる一群から選択されたアミノ酸配列と、(b)SEQ ID NO: 1-21からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c)SEQ ID NO: 1-21からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d)SEQ

Q I D N O : 1 - 2 1 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドの生産方法を提供する。この方法は、(a) このポリペプチドの発現に好適な条件下で、この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) このように発現したポリペプチドを回収するステップとを含む。

【 0 0 2 2 】

更に、本発明は、(a) S E Q I D N O : 1 - 2 1 からなる一群から選択されたアミノ酸配列と、(b) S E Q I D N O : 1 - 2 1 からなる一群から選択されたアミノ酸配列と 9 0 % 以上の配列同一性を有する天然のアミノ酸配列と、(c) S E Q I D N O : 1 - 2 1 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) S E Q I D N O : 1 - 2 1 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドに特異的に結合する単離された抗体を提供する。

10

【 0 0 2 3 】

更に、本発明は、(a) S E Q I D N O : 2 2 - 4 2 からなる一群から選択されたポリヌクレオチド配列と、(b) S E Q I D N O : 2 2 - 4 2 からなる一群から選択されたポリヌクレオチド配列と 9 0 % 以上の配列同一性を有する天然のポリヌクレオチド配列と、(c) 前記(a) に相補的なポリヌクレオチド配列と、(d) 前記(b) に相補的なポリヌクレオチド配列と、(e) 前記(a) 乃至(d) の R N A 等化物とで構成される一群から選択されたポリヌクレオチド配列を含む単離されたポリヌクレオチドを提供する。別法では、このポリヌクレオチドは、少なくとも 6 0 個の連続するヌクレオチドを含む。

20

【 0 0 2 4 】

更に本発明は、(a) S E Q I D N O : 2 2 - 4 2 からなる一群から選択されたポリヌクレオチド配列と、(b) S E Q I D N O : 2 2 - 4 2 からなる一群から選択されたポリヌクレオチド配列と 9 0 % 以上の配列同一性を有する天然のポリヌクレオチド配列と、(c) 前記(a) に相補的なポリヌクレオチド配列と、(d) 前記(b) に相補的なポリヌクレオチド配列と、(e) 前記(a) 乃至(d) の R N A 等化物とで構成される一群から選択されたポリヌクレオチド配列を含むポリヌクレオチド配列を有する標的ポリヌクレオチドをサンプルにおいて検出する方法を提供する。この方法は、(a) 前記サンプル内の標的ポリヌクレオチドと相補的な配列を構成する少なくとも 2 0 個の連続するヌクレオチドを含むプローブと前記サンプルをハイブリダイズさせるステップであって、前記プローブと前記標的ポリヌクレオチドまたはその断片とでハイブリダイゼーション複合体が形成される条件下で、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、該ステップと、(b) 前記ハイブリダイゼーション複合体の存在するか否かを検出し、存在する場合には随意選択でその収量を測定するステップとを含む。別法では、前記プローブは、少なくとも 6 0 個の連続するヌクレオチドを含む。

30

【 0 0 2 5 】

更に本発明は、(a) S E Q I D N O : 2 2 - 4 2 からなる一群から選択されたポリヌクレオチド配列と、(b) S E Q I D N O : 2 2 - 4 2 からなる一群から選択されたポリヌクレオチド配列と 9 0 % 以上の配列同一性を有する天然のポリヌクレオチド配列と、(c) 前記(a) に相補的なポリヌクレオチド配列と、(d) 前記(b) に相補的なポリヌクレオチド配列と、(e) 前記(a) 乃至(d) の R N A 等化物とで構成される一群から選択されたポリヌクレオチド配列を含むポリヌクレオチド配列を有する標的ポリヌクレオチドをサンプルにおいて検出する方法を提供する。この方法は、(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ステップと、(b) 増幅された前記標的ポリヌクレオチドまたはその断片が存在するか否かを検出し、存在する場合には随意選択でその収量を測定するステップとを含む。

40

【 0 0 2 6 】

更に本発明は、(a) S E Q I D N O : 1 - 2 1 からなる一群から選択されたアミノ酸

50

配列と、(b) SEQ ID NO: 1 - 21 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1 - 21 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1 - 21 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含む効果的な量のポリペプチド及び好適な医薬用賦形剤を含む組成物を提供する。一実施例では、SEQ ID NO: 1 - 21 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GCREC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。

【0027】

更に本発明は、(a) SEQ ID NO: 1 - 21 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1 - 21 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1 - 21 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1 - 21 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドのアゴニストとして効果的な化合物をスクリーニングする方法を提供する。この方法は、(a) この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) このサンプルのアゴニスト活性を検出するステップとを含む。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GCREC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。

10

20

【0028】

更に、本発明は、(a) SEQ ID NO: 1 - 21 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1 - 21 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1 - 21 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1 - 21 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドのアンタゴニストとして効果的な化合物をスクリーニングする方法を提供する。この方法は、(a) この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) このサンプルのアンタゴニスト活性を検出するステップとを含む。別法では、本発明は、この方法によって同定されたアンタ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GCREC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。

30

【0029】

更に本発明は、(a) SEQ ID NO: 1 - 21 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1 - 21 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1 - 21 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1 - 21 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドに特異的に結合する化合物をスクリーニングする方法を提供する。この方法は、(a) このポリペプチドを好適な条件下で少なくとも1つの化合物と結合させるステップと、(b) このポリペプチドとこの試験化合物との結合を検出して、このポリペプチドと特異的に結合する化合物を同定するステップとを含む。

40

【0030】

更に本発明は、(a) SEQ ID NO: 1 - 21 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1 - 21 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1 -

50

21からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1-21とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列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。このスクリーニング方法は、(a)このポリペプチドを、その活性が許容される条件下で少なくとも1つの化合物と結合させるステップと、(b)この試験化合物の存在下でのこのポリペプチドの活性を評価するステップと、(c)この試験化合物の存在下でのこのポリペプチドの活性と、この試験化合物の不在下でのこのポリペプチドの活性とを比較するステップとを含み、この試験化合物の存在下でのこのポリペプチドの活性の変化が、このポリペプチドの活性を調節する化合物の存在を示唆するという特徴を有する。

10

【0031】

更に本発明は、SEQ ID NO: 22-42からなる一群から選択された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)この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)この標的ポリヌクレオチドの発現の変化を検出するステップとを含む、該スクリーニング方法を提供する。

【0032】

本発明はさらに、試験化合物の毒性を評価する方法を提供する。この方法は、(a)核酸を含む生体サンプルを前記試験化合物で処理するステップと、(b)処理した前記生体サンプルの核酸をプローブとハイブリダイズするステップと、(c)ハイブリダイゼーション複合体の収量を測定するステップと、(d)前記処理した生体サンプルにおけるハイブリダイゼーション複合体の収量を、未処理の生体サンプルにおけるハイブリダイゼーション複合体の収量とを比較するステップとを含み、前記処理した生体サンプルにおけるハイブリダイゼーション複合体の収量の差異が試験化合物の毒性を示唆する。この方法における前記プローブは、(1)SEQ ID NO: 22-42からなる一群から選択されたポリヌクレオチド配列と、(2)SEQ ID NO: 22-42からなる一群から選択されたポリヌクレオチド配列と少なくとも90%の配列同一性を有する天然のポリヌクレオチド配列と、(3)前記(1)に相補的なポリヌクレオチド配列と、(4)前記(2)に相補的なポリヌクレオチド配列と、(5)前記(1)乃至(4)のRNA等価物とで構成される一群から選択されたポリヌクレオチド配列を含むポリヌクレオチドの連続する少なくとも20個のヌクレオチドを含む。また、前記ハイブリダイゼーションは、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る。また、前記標的ポリヌクレオチドが、(1)SEQ ID NO: 22-42からなる一群から選択されたポリヌクレオチド配列と、(2)SEQ ID NO: 22-42からなる一群から選択されたポリヌクレオチド配列と少なくとも90%の配列同一性を有する天然のポリヌクレオチド配列と、(3)前記(1)に相補的なポリヌクレオチド配列と、(4)前記(2)に相補的なポリヌクレオチド配列と、(5)前記(1)乃至(5)のRNA等価物とを含む。代替的に前記標的ポリヌクレオチドは前記ポリヌクレオチド配列の断片である。

20

30

【0033】

(本発明の記載について)

本発明のタンパク質及び核酸配列、方法について説明する前に、本発明は、ここに開示した特定の装置及び材料、方法に限定されず、その実施形態を変更できることを理解されたい。また、ここで用いられる用語は、特定の実施例のみを説明する目的で用いられたものであり、後述の請求の範囲によってのみ限定され、本発明の範囲を限定することを意図したものではないということも理解されたい。

40

【0034】

本明細書及び請求の範囲において単数形を表す「或る」、「その(この等)」は、文脈で明確に示していない場合は複数形を含むことに注意されたい。従って、例えば「或る宿主細胞」は複数の宿主細胞を含み、その「抗体」は複数の抗体は含まれ、当業者には周知の

50

等価物なども含まれる。

【0035】

本明細書で用いた全ての科学技術用語は、別の方法で定義されていない限り、本発明の属する技術分野の一般的な技術者が普通に解釈する意味と同じである。本明細書で記述したものと類似、或いは同等の全ての装置及び材料、方法は本発明の実施及びテストに使用できるが、好適な装置及び材料、方法をここに記す。本明細書に記載の全ての文献は、本発明に関連して使用する可能性のある文献に記載された細胞系、プロトコル、試薬、ベクターを記述し開示するために引用した。従来の発明を引用したからと言って、本発明の新規性が損なわれると解釈されるものではない。

【0036】

(定義)

用語「GCREC」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GCRECのアミノ酸配列を指す。

【0037】

用語「アゴニスト」は、GCRECの生物学的活性を強める、或いは模倣する分子を指す。このアゴニストは、GCRECに直接相互作用するか、或いはGCRECが関与する生物学的経路の成分と作用して、GCREC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。

【0038】

用語「アレル変異配列」は、GCRECをコードする遺伝子の別の形を指す。アレル変異配列は、核酸配列における少なくとも1つの変異によって生じ、変異mRNA若しくは変異ポリペプチドになり、これらの構造や機能は変わる場合もあれば変わらない場合もある。ある遺伝子は、天然型のアレル変異配列が存在しないもの、1つ或いは多数存在するものがある。一般にアレル変異配列を生じる変異は、ヌクレオチドの自然な欠失、付加、或いは置換による。これらの各変異は、単独或いは他の変異と同時に起こり、所定の配列内で一回或いはそれ以上生じる。

【0039】

GCREC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、GCRECと同じポリペプチド或いはGCREC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、GCREC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アレル変異配列との不相当或いは予期しないハイブリダイゼーション、並びにGCREC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GCREC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GCREC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、負に荷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含まれ、正に荷電したアミノ酸にはリシン及びアルギニンが含まれ得る。類似の親水性の値をもち極性非荷電側鎖を有するアミノ酸には、アスパラギン、グルタミン、セリン、トレオニンが含まれ得る。類似の親水性の値をもち非荷電側鎖を有するアミノ酸には、ロイシン、イソロイシン、バリン、グリシン、アラニン、フェニルアラニン及びチロシンが含まれ得る。

【0040】

用語「アミノ酸」及び「アミノ酸配列」は、オリゴペプチド、ペプチド、ポリペプチド、タンパク質配列、或いはそれらの任意の断片を指し、天然の分子及び合成分子を含む。「アミノ酸配列」が天然のタンパク質分子の配列を指す場合、「アミノ酸配列」及び類似の用語は、アミノ酸配列に記載したタンパク質分子に関連する完全で元のままのアミノ酸配列に限定するものではない。

10

20

30

40

50

【0041】

用語「増幅」は、核酸配列の複製物を作製することに関連する。一般に増幅は、この技術分野で周知のポリメラーゼ連鎖反応（PCR）技術によって行われる。

【0042】

用語「アンタゴニスト」は、GCRC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、GCRCに直接相互作用するか、或いはGCRCが関与する生物学的経路の成分と作用して、GCRC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。

【0043】

用語「抗体」は、抗原決定基と結合可能なFab及びF(ab')₂、及びそれらの断片、Fv断片などの無傷の分子を指す。GCRC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（例えば、マウス、ラット、若しくはウサギ）を免疫化するのに使用されるポリペプチド或いはオリゴペプチドは、RNAの翻訳から、或いは化学的に合成可能であり、必要に応じて担体タンパク質と結合させることも可能である。ペプチドと化学的に結合した一般に用いられる担体は、ウシ血清アルブミン、チログロブリン、及びキーホールリンペットヘモニアン（KLH）を含む。次ぎに、この結合したペプチドを用いて動物を免疫化する。

【0044】

用語「抗原決定基」は、特定の抗体と接触する分子の領域（即ちエピトープ）を指す。タンパク質或いはタンパク質の断片が、宿主動物を免疫化するのに用いられるとき、このタンパク質の種々の領域は、抗原決定基（タンパク質上の特定の領域或いは三次元構造体）に特異的に結合する抗体の産生を誘発し得る。抗原決定基は、抗体と結合するために無傷の抗原（即ち、免疫応答を引き出すために用いられる免疫原）と競合し得る。

【0045】

本明細書において「アンチセンス」は、特定の核酸配列のセンス（コーディング）鎖と塩基対を形成し得る任意の組成物を指す。アンチセンス成分には、DNAと、RNAと、ペプチド核酸（PNA）と、ホスホロチオネートやメチルホスホネート、ベンジルホスホネート（benzylphosphate）などの修飾された骨格（backbone linkage）を有するオリゴヌクレオチドと、2'-メトキシエチル糖または2'-メトキシエトキシ糖などの修飾された糖を有するオリゴヌクレオチドと、5-メチルシトシンまたは2'-deoxyuracil、7-deaza-2'-deoxyguanosineなどの修飾された塩基を有するオリゴヌクレオチドを含み得る。アンチセンス分子は、化学合成や転写を含む任意の方法で作ることができる。相補的アンチセンス分子は、一度細胞に導入されると、細胞によって作られた天然の核酸配列と塩基対となって二重鎖を形成し、転写や翻訳を阻害する。「負」または「マイナス」という表現はアンチセンス鎖であり、「正」または「プラス」という表現はセンス鎖である。

【0046】

用語「生物学的に活性」は、天然分子の構造的、調節的、或いは生化学的な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GCRC、合成のGCRC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。

【0047】

用語「相補的」は、塩基対合によってアニールする2つの一本鎖核酸配列間の関係を指す。例えば、配列「5'AGT3'」が相補的な配列「3'TCA5'」と対をなす。

【0048】

「所定のポリヌクレオチド配列を含む組成物」または「所定のアミノ酸配列を含む組成物」は広い意味で、所定のヌクレオチド配列若しくはアミノ酸配列を含む任意の組成物を指す。この組成物は、乾燥した製剤或いは水溶液を含み得る。GCRC若しくはGCRC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、プローブは、塩（例えば、NaCl）及び界面活性剤（例えば、SDS：ドデシル硫酸ナトリウム）、その他の物質（例えば、デンハート液、乾燥ミルク、サケ精子DNAなど）を含む水溶液に展開され得る。

【0049】

「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。伸長及び重複の両方によって構築されるコンセンサス配列もある。

10

【0050】

用語「保存的なアミノ酸置換」は、元のタンパク質の特性を殆ど変えない置換を指す。即ち、置換によってそのタンパク質の構造や機能が大きくは変わらず、そのタンパク質の構造、特にその機能が保存される。以下に、あるタンパク質の元のアミノ酸が別のアミノ酸に置換される保存的なアミノ酸置換を示す。

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

20

30

一般に、保存されたアミノ酸置換の場合は、a)置換された領域のポリペプチドの骨格構造、例えば、シートやヘリックス高次構造、b)置換された部位の分子の電荷または疎水性、及び/または、c)側鎖の大半が維持される。

40

【0051】

用語「欠失」は、1個以上のアミノ酸残基が欠如するアミノ酸配列の変化、或いは1個以上のヌクレオチドが欠如する核酸配列の変化を指す。

【0052】

用語「誘導体」は、化学修飾されたポリヌクレオチドまたはポリペプチドを指す。ポリヌクレオチド配列の化学修飾には、例えば、アルキル基、アシル基、ヒドロキシル基、或いはアミノ基による水素の置換がある。誘導体ポリヌクレオチドは、自然分子（未修飾の分子）の生物学的或いは免疫学的機能の少なくとも1つを維持するポリペプチドをコードする。誘導体ポリペプチドとは、もとのポリペプチドの生物学的機能、或いは免疫学的機能

50

の少なくとも1つを維持する、グリコシル化、ポリエチレングリコール化、或いは任意の同様のプロセスによって修飾されたポリペプチドのことである。

【0053】

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

【0054】

用語「断片」は、G C R E CまたはG C R E C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」の最大の長さは、親配列から1つのヌクレオチド/アミノ酸残基を差し引いた長さである。例えば、ある断片は、5 ~ 1000個の連続するヌクレオチド或いはアミノ酸残基を含む。プローブ、プライマー、抗原、治療用分子、またはその他の目的に用いる断片は、少なくとも5、10、15、16、20、25、30、40、50、60、75、100、150、250若しくは500個の連続するヌクレオチド或いはアミノ酸残基の長さである。断片は、優先的に分子の特定の領域から選択される場合もある。例えば、ポリペプチド断片は、所定の配列に示された最初の250若しくは500のアミノ酸（或いは、ポリペプチドの最初の25%または50%）から選択された連続するアミノ酸の所定の長さを含み得る。これらの長さは一例であり、配列表及び表、図面を含む明細書に記載の任意の長さが、本発明の実施例に含まれ得る。

10

【0055】

S E Q I D N O : 2 2 - 4 2 の断片は、例えば、この断片を得たゲノム内の他の配列とは異なる、S E Q I D N O : 2 2 - 4 2 を明確に同定する固有のポリヌクレオチド配列の領域を含む。S E Q I D N O : 2 2 - 4 2 のある断片は、例えば、ハイブリダイゼーションや増幅技術、またはS E Q I D N O : 2 2 - 4 2 を関連ポリヌクレオチド配列から区別する類似の方法に有用である。ある断片と一致するS E Q I D N O : 2 2 - 4 2 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。

20

【0056】

S E Q I D N O : 1 - 2 1 のある断片は、S E Q I D N O : 2 2 - 4 2 のある断片によってコードされる。S E Q I D N O : 1 - 2 1 のある断片は、S E Q I D N O : 1 - 2 1 を特異的に同定する固有のアミノ酸配列の領域を含む。例えば、S E Q I D N O : 1 - 2 1 のある断片は、S E Q I D N O : 1 - 2 1 を特異的に認識する抗体の開発における免疫原性ペプチドとして有用である。ある断片と一致するS E Q I D N O : 1 - 2 1 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決定できる。

30

【0057】

「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）、それに続く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。

【0058】

「相同性」は、2つ以上のポリヌクレオチド配列間または2つ以上のポリペプチド配列間の配列類似性である。この配列類似性は配列同一性と言い換えることができる。

40

【0059】

ポリヌクレオチド配列についての用語「パーセントの同一性」または「%の同一性」とは、標準化されたアルゴリズムを用いてアラインメントされる、2つ以上のポリヌクレオチド配列間の一致する残基の百分率のことである。このようなアルゴリズムは、標準化され再現できる方法で、2つの配列間のアラインメントを最適化するべく、配列にギャップを挿入して、より意味をもつ2つの配列間の比較を行うことができる。

【0060】

ポリヌクレオチド配列間の同一性のパーセントは、M E G A L I G N v e r s i o n 3 . 1 2 e 配列アラインメントプログラムに組込まれるC L U S T A L V アルゴリズムの

50

デフォルトパラメータを用いて決定可能である。このプログラムはLASERGENEソフトウェアパッケージの一部であり、分子生物学分析プログラム一式(DNA STAR, Madison WI)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。ポリヌクレオチド配列の対のアライメントの場合、デフォルトパラメータは、Ktuple = 2、gap penalty = 5、window = 4、「diagonals saved」= 4と設定する。「重み付けされた」残基重み付け表が、デフォルトとして選択された。同一性のパーセントは、アラインメントされたポリヌクレオチド配列間の「類似性のパーセント」としてCLUSTAL Vによって報告される。

10

【0061】

別法では、National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S. F. 他 (1990) J. Mol. Biol. 215:403-410)が提供する、広く用いられている無料の配列比較アルゴリズム一式が、NCBI (Bethesda, MD)を含む幾つかのソース及びインターネット (<http://WWW.ncbi.nlm.nih.gov/BLAST/>)で入手可能である。このBLASTソフトウェア一式には、既知のポリヌクレオチド配列と様々なデータベースの別のポリヌクレオチド配列とのアラインメントに用いられる「blastn」を含む、様々な配列分析プログラムが含まれる。「BLAST 2 Sequences」と呼ばれるツールが入手可能であり、2つのヌクレオチド配列の対を直接比較するために用いられる。「BLAST 2 Sequences」は、<http://WWW.ncbi.nlm.nih.gov/gorf/b12.html>にアクセスして、対話形式で利用できる。「BLAST 2 Sequences」ツールは、blastn 及び blastp (以下に記載)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、デフォルトを設定するギャップ及び他のパラメータと共に用いられる。例えば、2つのヌクレオチド配列を比較する場合、ある者は、デフォルトパラメータに設定された「BLAST 2 Sequences」ツール Version 2.0.12 (April-21-2000)でblastnを使用するであろう。そのようなデフォルトパラメータは、例えば、以下のようにする。

20

30

【0062】

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

同一性のパーセントは、例えば、特定の配列番号で決められた、所定の配列の全長に対して測定してもよいし、それより短い長さに対して、例えば、ある大きな所定の配列から得られた断片、例えば、連続する少なくとも、20または30、40、50、70、100、200のヌクレオチドの断片の長さに対して測定してもよい。このような長さは単なる例であり、配列表及び表、図面を含む明細書に記載の配列の任意の長さの断片を用いて、同一性のパーセントが測定される長さを示すことができる。

40

【0063】

高い同一性を示さない核酸配列でも、遺伝子コードの縮重によって類似のアミノ酸配列をコードし得る。縮重を利用して核酸配列を変え、それぞれが実質的に同じタンパク質をコードする様々な核酸配列を作製できることを理解されたい。

【0064】

50

ポリペプチド配列に用いられる用語「パーセントの同一性」または「%の同一性」とは、標準化されたアルゴリズムを用いてアラインメントされる2つ以上のポリペプチド配列間の一致する残基の百分率のことである。ポリペプチド配列アラインメントの方法は周知である。アラインメント方法の中には、保存的なアミノ酸置換を考慮したものもある。詳細に上述したこのような保存的な置換は、一般に、置換部位の電荷や疎水性が保存され、ポリペプチドの構造（従って機能も）が保存される。

【0065】

ポリペプチド配列間の同一性のパーセントは、MEGALIGNバージョン3.12e配列アラインメントプログラム（上記）に組込まれるCLUSTAL Vアルゴリズムのデフォルトパラメータを用いて決定可能である。CLUSTAL Vを用いる対方式のポリペプチド配列のアライメントの場合、デフォルトパラメータは、Ktuple = 1、gap penalty = 3、window = 5、及び「diagonals saved」= 5と設定する。PAM250マトリクスが、デフォルトの残基重み付け表として選択される。ポリヌクレオチドアラインメントと同様に、アラインメントされたポリペプチド配列の対の同一性のパーセントは、「類似性のパーセント」としてCLUSTAL Vによって報告される。

10

【0066】

別法では、NCBI BLASTソフトウェア一式が用いられる。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。そのようなデフォルトパラメータは、例えば、以下のようにする。

20

【0067】

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

同一性のパーセントは、例えば、特定の配列番号で決められた、所定のポリペプチド配列の全長に対して測定してもよいし、それより短い長さに対して、例えば、ある大きな所定のポリペプチド配列から得られた断片、例えば、連続する少なくとも15、20または30、40、50、70、150の残基の断片の長さに対して測定してもよい。このような長さは単なる例であり、配列表及び表、図面を含む明細書に記載の配列の任意の長さの断片を用いて、同一性のパーセントが測定される長さを示すことができる。

30

【0068】

「ヒト人工染色体（HAC）」は、約6 kb（キロベース）～10 MbのサイズのDNA配列を含み得る、安定した有糸分裂染色体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。

【0069】

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

40

【0070】

「ハイブリダイゼーション」とは、所定のハイブリダイゼーション条件下で、ある一本鎖ポリヌクレオチドがある相補的な一本鎖と塩基対を形成するアニーリングのプロセスである。特異的なハイブリダイゼーションとは、2つの核酸配列が高い相同性を有することを意味する。アニーリングが許容される条件下で、特異的なハイブリダイゼーション複合体が形成され、洗浄過程の後もハイブリダイズしたままである。洗浄過程は、ハイブリダイゼーションプロセスの厳密性即ちストリンジェンシー（stringency）の決定において特に重要であり、よりストリンジェントな条件では、非特異的な結合、即ち完全に

50

は一致しない核酸鎖間の対の結合が減少する。核酸配列間のアニーリングが許容される条件は、当業者によって日常的に決定され、ハイブリダイゼーションの間は一定であるが、洗浄過程は、目的のストリンジェンシーにするためにその最中に条件の変更が可能であり、ハイブリダイゼーション特異性が得られる。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100 µg/mlのせん断して変性したサケ精子DNAが含まれる。

【0071】

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、洗浄過程を行う際の温度によっても左右される。この洗浄温度は通常、所定のイオン強度とpHにおける特定の配列の熱融点(T_m)より約5~20℃低く選択される。このT_mは、(所定のイオン強度とpHの下)標的の配列の50%が完全に一致するプローブと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の条件は、周知であり、Sambrook, J. 他による, 1989, Molecular Cloning: A Laboratory Manual, 第2版の1-3巻, Cold Spring Harbor Press, Plainview NY; 特に2巻の9章に記載されている。

10

【0072】

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の濃度は、約0.1%のSDSが存在の下、約0.1~2×SSCの範囲である。通常は、ブロッキング試薬を用いて非特異的なハイブリダイゼーションを阻止する。このようなブロッキング試薬には、例えば、約100~200 µg/mlの切断され変性したサケ精子DNAが含まれる。約35~50% v/vの濃度のホルムアミドなどの有機溶剤が、例えば、RNAとDNAのハイブリダイゼーションなどの特定の場合に用いることができる。これらの洗浄条件の有用な改変は、当業者には周知である。特に高いストリンジェントな条件でのハイブリダイゼーションは、ヌクレオチド間の進化における類似性を示唆し得る。このような類似性は、それらのヌクレオチド及びコードされたポリペプチドが類似の役割を果たしていることを強く示唆する。

20

【0073】

用語「ハイブリダイゼーション複合体」は、相補的な塩基間の水素結合によって、形成された2つの核酸配列の複合体を指す。ハイブリダイゼーション複合体は溶液中(例えば、C₀tまたはR₀t分析)で形成されるか、或いは溶液中の1つの核酸配列と固体の支持物(例えば、紙、膜、フィルタ、チップ、ピン、或いはスライドガラス、または細胞及びその核酸を固定する任意の適当な基板)に固定されたもう一つの核酸配列とで形成される。

30

【0074】

用語「挿入」或いは「付加」は、1個以上のアミノ酸残基或いはヌクレオチドがそれぞれ追加されるアミノ酸配列或いは核酸配列の変化を指す。

【0075】

「免疫応答」は、炎症性疾患及び外傷、免疫異常、感染症、遺伝病などに関連する症状を指す。これらの症状は、細胞系及び全身防衛系に影響を及ぼすサイトカイン及びケモカイン、別の情報伝達分子などの様々な因子の発現という特徴をもつ。

40

【0076】

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GPCR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GCRCのポリペプチド断片またはオリゴペプチド断片を含む。

【0077】

用語「マイクロアレイ」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。

50

【0078】

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

【0079】

用語「調節」は、G C R E Cの活性の変化を指す。例えば、調節によって、G C R E C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。

【0080】

用語「核酸」及び「核酸配列」は、ヌクレオチド、オリゴヌクレオチド、ポリヌクレオチド、或いはそれらの断片を指し、一本鎖若しくは二本鎖であって、センス鎖或いはアンチセンス鎖であるゲノム起源若しくは合成起源のDNA或いはRNA、ペプチド核酸(PNA)、任意のDNA様物質、及びRNA様物質である。

10

【0081】

「機能的に結合した」は、第1の核酸配列と第2の核酸配列が機能的な関係にある状態を指す。例えば、プロモーターがコード配列の転写または発現に影響を与える場合、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。一般に、機能的に結合したDNA配列は、同じ読み枠内で2つのタンパク質をコードする領域が結合する必要がある場合は、非常に近接或いは連続する。

【0082】

「ペプチド核酸(PNA)」は、末端がリシンで終わるアミノ酸残基のペプチド骨格に結合した、少なくとも約5ヌクレオチドの長さのオリゴヌクレオチドを含む、アンチセンス分子または抗遺伝子剤を指す。この末端のリシンにより、この組成物が溶解性となる。PNAは、相補的な一本鎖DNAやRNAに優先的に結合して転写物の伸長を止め、ポリエチレングリコール化して細胞における寿命を延ばし得る。

20

【0083】

G C R E C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、G C R E Cの酵素環境に依存し、細胞の種類によって異なり得る。

【0084】

「プローブ」とは、同一配列或いはアレル核酸配列、関連する核酸配列の検出に用いる、G C R E C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、検出可能な標識またはレポーター分子が結合され単離されたオリゴヌクレオチドやポリヌクレオチドである。典型的な標識には、放射性アイソトープ及びリガンド、化学発光試薬、酵素がある。「プライマー」とは、相補的な塩基対を形成して標的のポリヌクレオチドにアニーリング可能な、通常はDNAオリゴヌクレオチドである短い核酸である。プライマーがポリヌクレオチドにアニーリングした後、あるDNAポリメラーゼ酵素によって、標的のDNA一本鎖に沿って伸長される。プライマーの組は、例えば、PCR法における核酸配列の増幅(及び同定)に用いることができる。

30

【0085】

本発明に用いられるプローブ及びプライマーは、既知の配列の少なくとも15の連続するヌクレオチドを含む。特異性を高めるために、より長いプローブ及びプライマーが用いることも可能である。例えば、開示した核酸配列の連続する少なくとも20または25、30、40、50、60、70、80、90、100、150のヌクレオチドを含む。プローブ及びプライマーは、上記した例より相当長いものも用いることができ、本明細書の表及び図面、配列表に示された任意の長さのヌクレオチドも用いることができることを理解されたい。

40

【0086】

プローブ及びプライマーの準備及び使用方法については、例えば、Sambrook, J. 他による、1989年、名称「Molecular Cloning: A Labo

50

ratory Manual」、第2版の1-3巻(Cold Spring Harbor Press, Plainview NY)、またはAusubel, F.M. 他による、1987年、名称「Current Protocols in Molecular Biology」(Greene Publi. Assoc. & Wiley-Intersciences, New York NY)、並びにInnis他による、1990年、名称「PCR Protocols, A Guide to Methods and Applications」(Academic Press, San Diego CA.)を参照されたい。PCR用のプライマーの組は、例えば、Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA)などのそのような目的のためのコンピュータプログラムを用いて、ある既知の配列から引き出すことができる。【0087】

プライマーとして用いるオリゴヌクレオチドは、当分野で周知のプライマー選択用のコンピュータプログラムで選択される。例えば、OLIGO 4.06ソフトウェアは、それぞれが最大100ヌクレオチドまでのPCR用のプライマーの対の選択、及び32,000塩基までの入力ポリヌクレオチド配列から最大5,000ヌクレオチドまでの大きなポリヌクレオチド及びオリゴヌクレオチドの分析に有用である。類似のプライマー選択用プログラムには、能力を拡大する追加の機能が含まれている。例えば、PrimOUプライマー選択プログラム(Genome Center at University of Texas South West Medical Center, Dallas TXより入手可能)は、メガベース配列から特定のプライマーを選択できるため、ゲノムワイドスコープ(genome-wide scope)におけるプライマーの設計に有用である。Primer3プライマー選択プログラム(Whitehead Institute/MIT Center for Genome Research, Cambridge MAより入手可能)によって、ユーザーは、プライマー結合部位として避けたい配列を指定できる「非プライミングライブラリ(mispriming library)」を入力できる。また、Primer3は、特にマイクロアレイのオリゴヌクレオチドの選択に有用である(後の方の2つのプライマー選択プログラムのソースコードは、それぞれのソースから得ることができ、ユーザーのニーズを満たすように変更することもできる)。PrimerGenプログラム(UK Human Genome Mapping Project Resource Centre, Cambridge UKより入手可能)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するため、アラインメントされた核酸配列の最も保存された領域或いは最も保存されていない領域のどちらかとハイブリダイズするプライマーを選択することができる。従って、このプログラムは、固有及び保存されたオリゴヌクレオチドやポリヌクレオチドの断片の同定に有用である。上記した任意の選択方法で同定されたオリゴヌクレオチドやポリヌクレオチドの断片は、例えば、PCR法やシーケンシングプライマー、マイクロアレイエレメント、或いはサンプルの核酸の完全或いは部分的に相補的なポリヌクレオチドを同定する特定のプローブなどの、ハイブリダイゼーション技術に有用である。オリゴヌクレオチドの選択方法は、上記した方法に制限されるものではない。【0088】

本明細書における「組換え核酸」は天然の配列ではなく、2つ以上の配列の離れたセグメントを人工的に組み合わせた配列である。この人工の組み合わせは、化学合成によって作られる場合も多いが、前出のSambrookに記載されたような遺伝子工学の技術を用いて核酸の離れたセグメントを人工的に操作する方がより一般的である。この「組換え核酸」には、単に核酸の一部の追加または置換、欠失によって変更された核酸も含む。組換え核酸は、あるプロモーター配列に機能的に結合した核酸配列を含む場合もある。このような組換え核酸は、例えば、ある細胞を形質転換するのに用いられるベクターの一部であり得る。【0089】

10

20

30

40

50

別法では、このような組換え核酸は、この組換え核酸を発現する哺乳動物のワクチン接種に用いると、その哺乳動物の防衛的な免疫応答を誘発する、ワクシニアウイルスに基づいたウイルスベクターの一部であり得る。

【0090】

「調節エレメント」は、通常は遺伝子の非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'の非翻訳領域(UTR)を含む。調節エレメントは、転写や翻訳、またはRNAの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。

【0091】

「レポーター分子」は、核酸、アミノ酸または抗体の標識に用いられる化学的または生化学的部分である。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、インヒビター、磁気粒子及びその他の当分野で既知の成分が含まれる。

10

【0092】

本明細書において、DNA配列に対する「RNA等価物」とは、基準となるDNA配列と同じ直鎖の核酸配列から構成されるが、窒素性塩基のチミンがウラシルに置換され、糖鎖の背骨がデオキシリボースではなくリボースからなる。

【0093】

用語「サンプル」は、その最も広い意味で用いられている。GCRC、GCRC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。

20

【0094】

用語「特異的結合」及び「特異的に結合する」は、タンパク質若しくはペプチドと、アゴニスト、抗体、アンタゴニスト、小分子、若しくは任意の天然若しくは合成の結合組成物との間の相互作用を指す。この相互作用は、結合する分子によって認識される、例えば、抗原決定基つまりエピトープなどのタンパク質の特定の構造の存在によって左右される。例えば、抗体がエピトープ「A」に対して特異的である場合、結合していない標識した「A」及び抗体を含む反応液に、エピトープAを含むポリペプチド或いは結合していない無標識の「A」が存在すると、抗体と結合する標識Aの量が減少する。

30

【0095】

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

【0096】

「置換」とは、一つ以上のアミノ酸またはヌクレオチドをそれぞれ別のアミノ酸またはヌクレオチドに置き換えることである。

【0097】

用語「基板」は、任意の好適な固体或いは半固体の支持物を指し、膜及びフィルタ、チップ、スライド、ウエハ、ファイバー、磁気または非磁気ビード、ゲル、チューブ、プレート、ポリマー、微小粒子、毛細管が含まれる。この基板には、壁または塹壕、ピン、チャンネル、細孔などの様々な表面形態があり、そこにポリヌクレオチドやポリペプチドが結合する。

40

【0098】

「転写イメージ」は、所定条件下での所定時間における特定の細胞の種類または組織による集合的遺伝子発現のパターンを指す。

【0099】

「形質転換」とは、外来DNAが受容細胞に導入されるプロセスのことである。形質転換

50

は、当分野で周知の種々の方法により、自然或いは人工の条件下で起こり、原核宿主細胞若しくは真核宿主細胞の中に外来核酸配列を挿入する任意の周知の方法によって行うことができる。この形質転換の方法は、形質転換される宿主細胞のタイプによって選択される。この方法には、バクテリオファージまたはウイルス感染、電気穿孔法（エレクトロポレーション）、リポフェクション、及び微粒子照射が含まれるが、これらに限定されるものではない。「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。

【0100】

本明細書における「遺伝子組換え生物」とは、当分野で周知の遺伝子組換え技術などを用いて、人間が生物の1つ以上の細胞に異種の核酸を導入した任意の生物であり、動物及び植物を含むが、それらに限定されるものではない。微量注入や組換えウイルスに感染させるなどの慎重な遺伝子操作によって、細胞の前駆体に直接或いは間接的に異種核酸を細胞に導入する。「遺伝子操作」とは、典型的な交雑育種や*in vitro*での受精ではなく、組換えDNA分子を導入することである。本発明に従った遺伝子組換え生物には、細菌及びラン藻類、菌類、植物、動物が含まれる。本発明の単離されたDNAは、当分野で周知の、例えば、感染、形質移入、形質転換、トランス接合（transconjugation）などの方法によって、宿主に導入することができる。本発明のDNAをそのような生物に導入する技術は周知であり、前出のSambrook他（1989）に記載されている。

【0101】

特定の核酸配列の「変異配列」とは、デフォルトパラメータ設定の「BLAST 2 Sequences」ツールVersion 2.0.9（May-07-1999）を用いるblastnによって、ある核酸配列のある長さに対する該特定の核酸配列の同一性が、少なくとも40%と決定された核酸配列のことである。このような核酸の対は、ある長さにおいて、例えば、少なくとも50%または60%、70%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、或いはそれ以上の同一性を示し得る。ある変異配列は、例えば、「アレル」変異配列（上述）または「スプライス」変異配列、「種」変異配列、「多型」変異配列と表すことができる。スプライス変異配列は基準分子と同一性が極めて高い可能性があるが、mRNAプロセッシング中のエキソンの択一的スプライシングによってポリヌクレオチドの数が多くなったり、少なくなったりする。対応するポリペプチドは、基準分子に存在する追加の機能ドメインを有したり、基準分子に存在するドメインが欠落したりし得る。種変異配列は、種によって異なるポリヌクレオチド配列である。得られるポリペプチドは、互いに高いアミノ酸同一性を有する。多型変異配列は、所定の種と種における特定の遺伝子のポリヌクレオチド配列が異なる。多型変異配列はまた、ポリヌクレオチド配列の1つのヌクレオチドが異なる「1ヌクレオチド多型」（SNP）も含み得る。SNPの存在は、例えば、或る集団、病態、病態の性向を示唆し得る。

【0102】

特定のポリペプチド配列の「変異体」とは、デフォルトパラメータ設定の「BLAST 2 Sequences」ツールVersion 2.0.9（May-07-1999）を用いるblastpによって、あるポリペプチド配列のある長さに対する該特定のポリペプチド配列の同一性が、少なくとも40%と決定されたポリペプチド配列のことである。このようなポリペプチドの対は、ある長さにおいて、例えば、少なくとも50%または60%、70%、80%、90%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、或いはそれ以上の同一性を示し得る。

【0103】

（発明）

本発明は、新規のヒトGタンパク質共役受容体（GCRC）及びGCRCをコードす

10

20

30

40

50

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増殖異常、神経の疾患、心血管疾患、胃腸疾患、自己免疫/炎症性の疾患、代謝障害、およびウイルス感染の診断、治療、及び予防に関する。

【0104】

表1は、本発明のポリヌクレオチド配列及びポリペプチド配列の識別番号を示す。各ポリヌクレオチドおよびそれに対応するポリペプチドは、1つのインサイトプロジェクト識別番号 (Incyte Project ID) に関連する。各ポリペプチド配列は、記載されているようにポリペプチド配列識別番号 (Polypeptide SEQ ID NO :) およびインサイトポリペプチド配列番号 (Incyte Polypeptide ID) の両方によって示されている。各ポリヌクレオチド配列は、記載されているようにポリヌクレオチド配列識別番号 (Polypeptide SEQ ID NO :) およびインサイトポリヌクレオチドコンセンサス配列番号 (Incyte Polypeptide ID) の両方によって示されている。

10

【0105】

表2は、GenBankタンパク質 (genept) データベースにおいてBLAST解析により同定された本発明のポリペプチドに相同性を有する配列を示す。列1および列2はそれぞれ、本発明の各ポリペプチドに対するポリペプチド配列識別番号 (Polypeptide SEQ ID NO :) およびそれに対応するインサイトポリペプチド配列番号 (Incyte Polypeptide ID) を示す。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って本明細書の一部とする。

20

【0106】

表3は、本発明の各ポリペプチドの様々な構造的特徴を示す。列1および列2はそれぞれ、本発明の各ポリペプチドのポリペプチド配列識別番号 (SEQ ID NO :) およびそれに対応するインサイトポリペプチド配列番号 (Incyte Polypeptide 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、潜在的なリン酸化部位および潜在的なグリコシル化部位を示す。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。

30

【0107】

表2および表3は共に、本発明の各ポリペプチドの特性を要約したものであって、これらの特性は請求するポリペプチドがGタンパク質共役受容体であることを立証するものである。例えば、SEQ ID NO : 7は、残基M1からV306まで、ネズミ臭気物質レセプターMOR83 (GenBank ID g6178006) と85%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を参照)。BLASTの確率スコアは $5.5e^{-141}$ であり、探しているポリペプチド配列アラインメントが偶然の一致により得られる確率を示す。SEQ ID NO : 7はまた、7膜内外受容体 (ロドプシンファミリー) 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を参照)。BLIMPS、MOTIFS、およびPROFILESCAN解析から得られたデータによって、SEQ ID NO : 7が嗅覚のGタンパク質共役受容体であることが裏付けられた。SEQ ID NO : 1、SEQ ID NO : 2、SEQ ID NO : 3、SEQ ID NO : 4、SEQ ID NO : 5、SEQ ID NO : 6、SEQ ID NO : 8、SEQ ID NO : 9、SEQ ID NO : 10、SEQ ID NO :

40

50

11、SEQ ID NO: 12、SEQ ID NO: 13、SEQ ID NO: 14、SEQ ID NO: 15、SEQ ID NO: 16、SEQ ID NO: 17、SEQ ID NO: 18、SEQ ID NO: 19、SEQ ID NO: 20、SEQ ID NO: 21も同様の方法で解析してアノテーションを付けた。SEQ ID NO: 1 - 21の解析のためのアルゴリズムおよびパラメータを表7に記載する。

【0108】

表4に示されているように、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、或いはこれらの2種類の配列のあらゆる組み合わせを用いて組み立てた。列1および列2はそれぞれ、本発明の各ポリヌクレオチドのポリヌクレオチド配列識別番号(Polynucleotide SEQ ID NO:)およびそれに対応するインサイトポリヌクレオチドコンセンサス配列番号(Insite Polynucleotide 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: 22 - 42を同定するため、或いはSEQ ID NO: 22 - 42と関連するポリヌクレオチド配列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から推定されるコード配列(エキソン)、および/またはcDNAおよびゲノムDNAの両方からなる群に対応する識別番号を示す。これらの配列を用いて本発明の完全長ポリヌクレオチド配列を組み立てた。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。

10

20

【0109】

表4の列5に示されている識別番号は、具体的には、例えばインサイトcDNAおよびそれらに対応するcDNAライブラリの識別番号を示す。例えば、6871486H1はインサイトcDNA配列の識別番号であり、BRAGNON02はそれが由来するcDNAライブラリの識別番号である。cDNAライブラリが示されていないインサイトcDNAは、プールされているcDNAライブラリ(例えば、70171099V1)に由来する。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、g5743982)の識別番号の場合もある。または、列5の識別番号は、Genscan分析によって推定されるゲノムDNAのコード領域の場合もある。例えば、GNN.g6671985_006は、Genscan推定コード配列の識別番号であって、g6671985がGenscan分析によって得られたGenBankの配列の識別番号である。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(実施例4を参照)。または、列5の識別番号は、“exon-stitching”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。例えば、FL2289894_00001は“stitched”配列を示し、ここで2289894はアルゴリズムが適用された配列のクラスタの識別番号であり、00001はアルゴリズムによって生成された予測数である。(実施例5を参照)。または、列5の識別番号は、“exon-stretching”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インサイト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、それに相当するインサイトcDNAの識別番号は示されていない。

30

40

【0110】

表5は、インサイトcDNA配列を用いて組み立てられたこれらの完全長ポリヌクレオチド配列が由来する代表的なcDNAライブラリを示す。代表的なcDNAライブラリとは、上記ポリヌクレオチド配列の組み立ておよび決定に用いられたインサイトcDNA配列を最も多く含むインサイトcDNAライブラリのことである。表5に示されているcDNAライブラリを作製するために用いた組織およびベクターが表6に示されている。

【0111】

本発明はまた、GCRECの変異体も含む。好適なGCRECの変異体は、GCRECの

50

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G C R E C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。

【0112】

本発明はまた、G C R E C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、G C R E CをコードするS E Q I D N O : 2 2 - 4 2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。配列表に示したS E Q I D N O : 2 2 - 4 2のポリヌクレオチド配列は、窒素系塩基のチミンがウラシルに置換され、糖鎖の背骨がデオキシリボースではなくリボースからなる等価R N A配列を含む。

【0113】

本発明はまた、G C R E C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、G C R E Cをコードするポリヌクレオチド配列と少なくとも70%のポリヌクレオチド配列同一性、或いは少なくとも85%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。本発明の特定の実施形態は、S E Q I D N O : 2 2 - 4 2からなる一群から選択された核酸配列と少なくとも70%のポリヌクレオチド配列同一性、或いは少なくとも85%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S E Q I D N O : 2 2 - 4 2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の変異配列を提供する。上記したポリヌクレオチド変異配列は何れも、G C R E Cの機能的或いは構造的特徴の少なくとも1つを有するアミノ酸配列をコードする。

【0114】

遺伝暗号の縮重により作り出され得るG C R E C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、可能なコドン選択に基づいた組み合わせの選択によって作り出され得る可能なポリヌクレオチド配列の変異の全てが含まれ得る。これらの組み合わせは、天然のG C R E C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。

【0115】

G C R E C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G C R E C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有するG C R E C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。特定のコドンが宿主によって利用される頻度に基づいてコドンを選択して、ペプチドの発現が特定の真核細胞または原核宿主に発生する割合を高めることが可能である。コードされたアミノ酸配列を変えないで、G C R E C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R N A転写物を作ることにある。

【0116】

本発明はまた、G C R E C及びその誘導体をコードするD N A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。作製後にこの合成配列を、当分野で良く知られた試薬を用いて、種々の入手可能な発現ベクター及び細胞系の何れの中にも挿入可能である。更に、合成化学を用いて、G C R E Cまたはその任意の断片をコードする配列の中に突然変異を導入することも可能である。

【0117】

更に本発明には、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、S E Q I D N O : 2 2 - 4 2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、W a h l , G . M . 及びS . L . B e r g e r (1 9 8 7) M e t h o d s E n z y m o l . 1 5 2 : 3 9 9 - 4 0 7 ; a

10

20

30

40

50

nd 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。

【0118】

当分野で周知のDNAのシーケンシング方法を用いて、本発明の何れの実施例も実行可能である。この方法には、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)、或いはELONGASE増幅システム (Life Technologies, Gaithersburg MD) にみられるような校正エキソヌクレアーゼとポリメラーゼとの組み合わせなどの酵素が用いられる。好ましくは、MICROLAB2200液体転移システム (Hamilton, Reno, NV)、PTC200 Thermal Cycler 200 (MJ Research, Watertown MA) 及びABI CATALYST 800 (PE Biosystems) などの装置を用いて配列の準備を自動化する。次に、ABI 373 或いは377 DNAシーケンシングシステム (PE Biosystems)、MEGABACE 1000 DNAシーケンシングシステム (Molecular Dynamics, Sunnyvale CA) または当分野で周知の他の方法を用いてシーケンシングを行う。得られた配列を当分野で周知の様々なアルゴリズムを用いて分析する (例えば、Ausubel, F. M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R. A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853. を参照)。

10

20

【0119】

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、GC REC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。例えば制限部位PCR法を利用する1つの方法では、一般的なプライマー及び入れ子プライマー (nested primer) 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 (例えば、Sarkar, G. (1993) *PCR Methods Applic* 2: 318-322 を参照)。逆PCR法を用いる別法では、広範な方向に伸長して環状化した鋳型から未知の配列を増幅するプライマーを用いる。この鋳型は、既知のゲノム遺伝子座及びその周辺の配列を含む制限断片に由来する (例えば、Triglia, T. (1988) *Nucleic Acids Res* 16: 8186 を参照)。キャプチャPCR法を用いる第3の方法は、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断片のPCR増幅を含む (例えば、Lagerstrom, M. 他 (1991) *PCR Methods Applic* 1: 111-119 を参照)。この方法では、多数の制限酵素による消化及びライゲーションを用いて、PCRを行う前に未知の配列の領域の中に組換え二本鎖配列を挿入することが可能である。また、当分野で周知の別の方法を用いて未知の配列を得ることも可能である。 (例えば、Parker, J. D. 他 (1991) *Nucleic Acids Res.* 19: 3055-3060 を参照)。更に、PCR、ネスト化プライマー、PROMOTER FINDER ライブラリ (Clontech, Palo Alto CA) を用いれば、ゲノムDNA内の歩行が可能である。この方法で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/エキソン接合部を探すのに有用である。全てのPCR法をベースにした方法では、プライマーは、市販のOLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) 或いは別の好適なプログラムなどを用いて、長さが22~30ヌクレオチド、GC含量が50%以上、約68~72の温度で鋳型に対してアニーリングするよう設計される。

30

40

50

【0120】

完全長のcDNAをスクリーニングする場合は、大きなcDNAを含むようにサイズが選択されたライブラリを用いるのが好ましい。更に、オリゴd(T)ライブラリが完全な長さのcDNAを産生できない場合は、遺伝子の5'領域を有する配列を含むものが多いランダムに初回抗原刺激を受けたライブラリが有用である。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。

【0121】

市販のキャピラリー電気泳動システムを用いて、シーケンシングまたはPCR産物のヌクレオチド配列のサイズの分析、または確認が可能である。詳しくは、キャピラリーシーケンシングには、電気泳動による分離のための流動性ポリマー、及び4つの異なるヌクレオチドに特異的なレーザーで活性化される蛍光色素、放出された波長の検出に利用するCCDカメラを使用することが可能である。出力/光強度は、適切なソフトウェア(例えば、GENOTYPER及びSEQUENCE NAVIGATOR、PE Biosystems)を用いて電気信号に変換され、サンプルのローディングからコンピュータ分析までのプロセス及び電子データ表示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場合もあるDNAの小片のシーケンシングに特に適している。

【0122】

本発明の別の実施例では、GCREC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GCREC、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GCRECのクローニング及び発現に利用可能である。

【0123】

種々の目的でGCREC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ニング、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。ランダムな断片によるDNAの混合や遺伝子断片と合成オリゴヌクレオチドのPCR再組み立てを用いて、ヌクレオチド配列の組換えが可能である。例えば、オリゴヌクレオチド媒介性定方向突然変異誘発を利用して、新しい制限部位を生成する突然変異の導入、グリコシル化パターンの変更、コドン選択の変更、スプライスバリアントの作製等が可能である。

【0124】

本発明のヌクレオチドを、MOLECULAR BREEDING (Maxygen Inc., Santa Clara CA; 米国特許第5,837,458号; Chang, C.-C. 他(1999) Nat. Biotechnol. 17:793-797; Christians, F.C. 他(1999) Nat. Biotechnol. 17:259-264; Cramer, A. 他(1996) Nat. Biotechnol. 14:315-319)などのDNAシャフリング技術を用いてシャフリングして、GCRECの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GCRECの生物学的特性を変更或いは改良することができる。DNAシャフリングは、PCR法による遺伝子断片の組換えで遺伝子変異体のライブラリを製作するプロセスである。次に、このライブラリを、目的の特性を有する遺伝子変異体を同定するために選択或いはスクリーニングする。これらの好ましい変異体をプールし、DNAシャフリング及び選択/スクリーニングを繰り返す。従って、人工的な育種及び急速な分子の進化によって多様な遺伝子が作られる。例えば、ランダムな位置に変異がある1つの遺伝子の断片を、目的の特性が最適化するまで、組換え及びスクリーニング、シャフリングを実施することもできる。別法では、所定の遺伝子の断片を、同じ或いは異なる種の間で同じ遺伝子ファミリーの相同な遺伝子の断片で組換え、それによってプロトコルに従った調節可能な方法で、多数の天然遺伝子の遺伝子多様性を最大にすることができる。

10

20

30

40

50

【0125】

別の実施例によれば、GCREC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) *Nucl. Acids Res. Symp. Ser. 7: 215-223*; 及びHorn, T.他(1980) *Nucl. Acids Res. Symp. Ser. 225-232*を参照)。別法として、化学的方法を用いてGCREC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、ペプチド合成は種々の固相技術を用いて実行可能である（例えば、Creighton, T. (1984) *Proteins. Structures and Molecular Properties*, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. 10
ら(1995) *Science 269: 202-204*を参照)。また、合成の自動化は例えばABI 431Aペプチドシンセサイザー(PE Biosystems)を用いて達成し得る。更にGCREC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。

【0126】

このペプチドは、分離用高速液体クロマトグラフィー（例えば、Chiez, R.M. 及び F.Z. Regnier (1990) *Methods Enzymol. 182: 392-421*を参照）を用いて実質的に精製可能である。合成されたペプチドの組成 20
は、アミノ酸分析或いはシーケンシングにより確認することができる（例えば、Creighton、前出、pp 28-53を参照）。

【0127】

生物学的に活性なGCRECを発現させるために、GCREC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GCREC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。このようなエレメントは、その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、GCRECをコードする配列のより効果的な翻訳を達成する 30
ことが可能である。このようなシグナルには、ATG開始コドン及びコザック配列などの近傍の配列が含まれる。GCREC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ATG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なければならない。外来性の翻訳要素及び開始コドンは、自然及び合成の様々なものから得ることが可能である。用いられる特定の宿主細胞系に好適なエンハンサーを含めることで発現の効率を高めることが可能である。（例えば、Scharf, D. 他(1994) *Results Probl. Cell Differ. 201-18-162*.を参照）。 40

【0128】

当業者に周知の方法を用いて、GCREC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, ch. 9章及び13章1-4章を参照）。 40

【0129】

種々の発現ベクター/宿主系を利用して、GCRECをコードする配列の保持及び発現が可能である。これらには、限定するものではないが、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れた細菌などの微生物や、酵母菌発現ベクターで形質転換された酵母菌や、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；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）で形質転換された植物細胞系や、動物細胞系などが含まれる（例えば、前出のSambrook、前出のAusubel、Van Heeke, G. および S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509、Engelhard, E.K. 他 (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227、Sandig, V. 他 (1996) *Hum. Gene Ther.* 7:1937-1945、Takamatsu, N. (1987) *EMBOJ.* 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、Harrington, J.J. 他 (1997) *Nat. Genet.* 15:345-355を参照)。レトロウイルス、アデノウイルス、ヘルペスウイルスまたはワクシニアウイルス由来の発現ベクター、または種々の細菌性プラスミド由来の発現ベクターを用いて、ヌクレオチド配列を標的器官、組織または細胞集団へ輸送することができる (Di Nicola, M. 他 (1998) *Cancer Gen. Ther.* 5(6):350-356、Yu, M. 他 (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344、Buller, R.M. 他 (1985) *Nature* 317(6040):813-815; McGregor, D.P. 他 (1994) *Mol. Immunol.* 31(3):219-226、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。

【0130】

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、GCREC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、GCREC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、PBLUESCRIPT (Stratagene, La Jolla CA) または pSPORT1 プラスミド (GIBCO BRL) 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GCRECをコードする配列をライゲーションすると lacZ 遺伝子が破壊され、組換え分子を含む形質転換された細菌の同定のための比色スクリーニング法が可能となる。更に、これらのベクターを用いて、クローニングされた配列の in vitro での転写、ジデオキシンスクリーニング、ヘルパーファージによる一本鎖の救出、入れ子状態の欠失を作り出すことが可能である (例えば、Van Heeke, G. 及び S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509. を参照)。例えば、抗体の産生のためなどに多量のGCRECが必要な場合は、GCREC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、強力に発現を誘発する T5 または T7 バクテリオファージプロモーターを含むベクターを使用できる。

【0131】

GCRECの発現に酵母の発現系の使用が可能である。因子やアルコールオキシダーゼやPGHプロモーターなどの構成型或いは誘導型のプロモーターを含む多種のベクターが、酵母菌サッカロミセス-セレビジエまたは Pichia pastoris 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、Ausubel, 1995, 前出、Bitter, G. A. ら (1987) *Methods Enzymol.* 153: 516-544、及び Scorer, C. A. ら (1994) *Bio/Technology* 121-181-184. を参照)。

【0132】

植物系も GCREC の発現に使用可能である。GCREC をコードする配列の転写は、例えば、CaMV 由来の 35S 及び 19S プロモーターなどのウイルスプロモーターが単独で、或いは TMV (例えば、Coruzzi, G. ら (1984) *EMBO J.* 3: 1671-1680; Broglie, R. ら (1984) *Science* 224: 838-843; および Winter, J. ら (1991) *Results Probl. Cell Differ.* 17: 85-105 を参照) 由来のオメガリーダー配列と組み合わせて促進される。これらの作製物は、直接の DNA 形質転換 10
或いは病原体を介したトランスフェクションによって、植物細胞の中に導入可能である。(例えば、The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill NY, pp. 191-196 を参照)。

【0133】

哺乳動物細胞では、多種のウイルスベースの発現系が利用され得る。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 GCREC をコードする配列を結合し得る。ウイルスの 20
ゲノムの非必須の E1 または E3 領域への挿入により、感染した宿主細胞に GCREC を発現する生ウイルスを得ることが可能である (Logan, J. 及び Shenk, T. (1984) *Proc. Natl. Acad. Sci.* 81: 3655-3659 を参照)。さらに、ラウス肉腫ウイルス (RSV) エンハンサーなどの転写エンハンサーを用いて、哺乳動物宿主細胞における発現を増大させることが可能である。タンパク質を高レベルで発現させるために、SV40 または EBV を基にしたベクターを用いることが可能である。

【0134】

ヒト人工染色体 (HAC) を用いて、プラスミドで発現しそれに含まれているものより大きな DNA の断片を供給可能である。治療のために約 6 kb ~ 10 Mb の HACs を作製 30
し、従来の輸送方法 (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)。

【0135】

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 GCREC 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、GCREC 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。ベクターの導入の後、細胞を選択培地に移す前に、強化培地で約 1 ~ 2 日の間増殖させる。選択マーカーの目的は選択的な媒介物に対する抵抗性を与えるとともに、その存在により導入された配列を確実に発現する細胞の増殖及び回収が可能となる 40
。安定的に形質転換された細胞の耐性クローンは、その細胞型に好適な組織培養技術を用いて増殖可能である。

【0136】

任意の数の選択系を用いて、形質転換された細胞系を回収することが可能である。選択系には、以下のものに限定はしないが、単純ヘルペスウイルスチミジンキナーゼ遺伝子及びアデニンホスホリボシルトランスフェラーゼ遺伝子が含まれ、それぞれ tk^r または ap^r 細胞において使用される。(例えば、Wigler, M. 他 (1977) *Cell* 11: 223-232; 及び Lowy, I. 他 (1980) *Cell* 22: 817-823 を参照)。また代謝拮抗物質、抗生物質或いは除草剤への耐性を選択のベース 50

として用いることができる。例えば dhfr はメトトレキセートに対する耐性を与え、neo はアミノグリコシッドネオマイシン及び G-418 に対する耐性を与え、als 或いは pat はクロルスルフロン (chlorosulfuron)、ホスフィノトリシンアセチルトランスフェラーゼ (phosphinotricin acetyltransferase) に対する耐性を与える (例えば、Wigler, M. 他. (1980) Proc. Natl. Acad. Sci. 77:3567-3570; Colbere-Garapin, F. 他 (1981) J. Mol. Biol. 150:1-14 を参照)。さらに選択に利用できる遺伝子、例えば、代謝のために細胞が必要なものを変える trpB 及び hisD が文献に記載されている (例えば、Hartman, S. C. 及び R. C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-51 を参照)。アノトシアニン、緑色蛍光タンパク質 (GFP; Clontech)、グルクロニダーゼ及びその基質 GUS, ルシフェラーゼ及びその基質ルシフェリンなどの可視マーカーが用いられる。緑色蛍光タンパク質 (GFP) (Clontech, Palo Alto, CA) も使用できる。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 (例えば、Rhodes, C. A. 他 (1995) Methods Mol. Biol. 55:121-131 を参照)。

10

【0137】

マーカー遺伝子の発現の存在 / 不在によって目的の遺伝子の存在が示されても、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、GCREC 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、GCREC 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 GCREC 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。

20

【0138】

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

30

【0139】

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 GCREC の発現の検出及び計測のための免疫学的方法は、当分野で周知である。このような技法には、酵素に結合したイムノソルベントアッセイ (ELISA)、ラジオイムノアッセイ (RIA)、蛍光標示式細胞分取器 (FACS) などがある。GCREC 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その他のアッセイは、当分野では十分に知られている。(例えば、Hampton, R. 他. (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN, Sect. IV; Coligan, J. E. 他 Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York, NY; 及び Pound, J. D. (1990) Immunochemical Protocols, Humana Press, Totowa NJ)。

40

【0140】

50

種々の標識技術及び結合技術が当業者には周知であり、様々な核酸アッセイおよびアミノ酸アッセイに用いられ得る。G C R E Cをコードするポリヌクレオチドに関連する配列を検出するための、標識されたハイブリダイゼーションプローブ或いはPCRプローブを生成する方法には、オリゴ標識化、ニックトランスレーション、末端標識化、または標識されたヌクレオチドを用いるPCR増幅が含まれる。別法として、G C R E Cをコードする配列、またはその任意の断片をmRNAプローブを生成するためのベクターにクローニングすることも可能である。当分野では周知であり市販されているこのようなベクターを、T7, T3, またはSP6などの好適なRNAポリメラーゼ及び標識されたヌクレオチドの追加によって、in vitroでのRNAプローブの合成に用いることができる。これらの方法は、例えば、Amersham Pharmacia Biotech及びPromega (Madison WI)、U.S. Biochemical Corp (Cleveland OH)が市販する種々のキットを用いて行うことができる。容易な検出の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、コファクター、インヒビター、磁気粒子、及び放射性核種、酵素、蛍光剤、化学発光剤、色素産生剤などが含まれる。

10

20

30

40

50

【0141】

G C R E C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。形質転換された細胞から産生されたタンパク質が分泌されるか細胞内に留まるかは、使用されるその配列及び/またはそのベクターによる。G C R E C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G C R E C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。

【0142】

更に、挿入した配列の発現調節能力または発現したタンパク質を所望の形にプロセッシングする能力によって宿主細胞株が選択される。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 (lipidation)、及びアシル化が含まれるが、これらに限定されるものではない。タンパク質の「プレプロ」または「プロ」形を切断する翻訳後のプロセッシングを利用して、標的タンパク質、折りたたみ及び/または活性を特定することが可能である。翻訳後の活性のための特定の細胞装置及び特徴のある機構をもつ種類の宿主細胞 (例えば、CHO、HeLa、MDCK、MEK293、WI38) がAmerican Type Culture Collection (ATCC; Bethesda, MD) より入手可能であり、外来のタンパク質の正しい修飾及びプロセッシングを確実にするために選択される。

【0143】

本発明の別の実施例では、G C R E C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G C R E Cタンパク質が、G C R E Cの活性のインヒビターに対するペプチドライブラリのスクリーニングを促進し得る。また、異種タンパク質部分及び異種ペプチド部分が、市販の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。このような部分には、グルタチオンSトランスフェラーゼ (GST)、マルトース結合タンパク質 (MBP)、チオレドキシン (Trx)、カルモジュリン結合ペプチド (CBP)、6-His、FLAG、c-mc、赤血球凝集素 (HA) が含まれるが、これらに限定されるものではない。GST及びMBP、Trx、CBP、6-Hisによって、固定されたグルタチオン、マルトース、フェニルアルシン酸化物 (phenylarsine oxide)、カルモジュリン、金属キレート樹脂のそれぞれで同族の融合タンパク質の精製が可能となる。FLAG、c-mc、及び赤血球凝集素 (HA) によって、これらのエピトープ標識を特異的に認識する市販のモノクローナル抗体及びポリクローナル抗体を用いた融合タンパク質の免疫親和性の精製ができる。また、G C R E C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、G C R E C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と精製の方法は、Ausubel. (1995、前出 ch 10) に記載されている。市販されている様々なキットを用いて、融合タンパク質の発現及び精製を促進できる。

【0144】

本発明の別の実施例では、TNTウサギ網状赤血球可溶化液またはコムギ胚芽抽出系 (Promega) を用いて *in vitro* で放射能標識した GCREC の合成が可能である。これらの系は、T7 または T3、SP6 プロモーターと機能的に結合したタンパク質をコードする配列の転写と翻訳をつなげる。翻訳は、例えば、³⁵S メチオニンである放射能標識されたアミノ酸前駆体の存在の下で起こる。

【0145】

本発明の GCREC またはその断片を用いて、GCREC 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、GCREC 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。

10

【0146】

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 GCREC 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 (Coligan, J. E. 他 (1991) Current Protocols in Immunology 1(2) の5章等を参照)。同様に、化合物は、GCREC 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 GCREC 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。GCREC を発現する細胞または GCREC を含有する細胞膜断片を試験化合物と接触させて、GCREC 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。

20

【0147】

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、フルオロフォア、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 GCREC と結合させるステップと、GCREC 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、細胞遊離剤、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、溶液中で遊離させるか固体支持体に固定させる。

30

【0148】

本発明の GCREC またはその断片を用いて、GCREC 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、GCREC が少なくとも1つの試験化合物と結合する、GCREC 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 GCREC の活性が試験化合物不在下での GCREC の活性と比較する。試験化合物の存在下での GCREC の活性の変化は、GCREC 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 GCREC の活性に適した条件下で GCREC を含む *in vitro* または細胞遊離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、GCREC 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。

40

【0149】

別の実施例では、胚性幹細胞 (ES細胞) における相同組換えを用いて動物モデル系内で

50

、GCREC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ES細胞は初期のマウス胚に由来し、培地で増殖させることができる。このES細胞は、ネオマイシンホスホトランスフェラーゼ遺伝子（neo:Capecchi, M.R. (1989) Science 244:1288-1292）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、Cre-loxP系を用いて相同組換えを行い、組織特異的または発生段階特異的に目的遺伝子をノックアウトする（Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. 他 (1997) Nucleic Acids Res. 25:4323-4330）。形質転換したES細胞を同定し、例えばC57BL/6マウス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遺伝子組換え動物は、潜在的な治療薬や毒性薬剤で検査することができる。

【0150】

GCRECをコードするポリヌクレオチドを *in vitro* でヒト胚盤胞由来のES細胞において操作することが可能である。ヒトES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（Thomson, J.A. 他 (1998) Science 282:1145-1147）。

【0151】

GCREC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、GCRECをコードするポリヌクレオチドの或る領域を動物ES細胞に注入し、注入した配列を動物細胞ゲノムに組み込ませる。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、潜在的な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GCRECを乳汁内に分泌するなどGCREC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（Janne, J. 他 (1998) Biotechnol. Annu. Rev. 4:55-74）。

【0152】

（治療）

GCRECのある領域とGタンパク質共役受容体のある領域との間に、例えば配列及びモチーフの文脈における化学的及び構造的類似性が存在する。加えて、GCRECの発現は、卵巣腫瘍、前立腺、白血球、小脳、及び脳組織に密接に関係している。従って、GCRECは、増殖異常、神経の疾患、心血管疾患、胃腸疾患、自己免疫/炎症性の疾患、代謝障害、およびウイルス感染においてある役割を果たすと考えられる。GCREC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、GCRECの発現または活性を低下させることが望ましい。また、GCREC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、GCRECの発現または活性を増大させることが望ましい。

【0153】

従って、一実施例において、GCREC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GCREC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、このような疾患には増殖異常、神経の疾患、心血管疾患、胃腸疾患、自己免疫/炎症性の疾患、代謝障害、およびウイルス感染が含まれ、増殖異常の中には日光性角化症及びアテローム性動脈硬化、滑液包炎、硬変、肝炎、混合型結合組織病（MCTD）、骨髄線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症、並びに腺癌及び白血病、リンパ腫、黒色腫、骨髄腫、肉腫、及び奇形癌、

具体的には、副腎、膀胱、骨、骨髄、脳、乳房、頸部、胆嚢、神経節、消化管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、子宮の癌などが含まれ、神経の疾患の中には、癲癇、虚血性脳血管障害、脳卒中、大脳新生物、アルツハイマー病、ピック病、ハンチントン病、痴呆、パーキンソン病及びその他の錐体外路障害、筋萎縮性側索硬化及びその他の運動ニューロン障害、進行性神経性筋萎縮症、色素性網膜炎、遺伝性運動失調、多発性硬化症及び他の脱髄疾患、細菌性及びウイルス性髄膜炎、脳膿瘍、硬膜下蓄膿症、硬膜外膿瘍、化膿性頭蓋内血栓性静脈炎、脊髄炎及び神経根炎、ウイルス性中枢神経系疾患と、クールー及びクロイツフェルトヤコブ病、ゲルストマン症候群、Gerstmann-Straussler-Scheinker症候群を含むプリオン病と、致死性家族性不眠症、神経系性栄養病及び代謝病、神経線維腫症、結節硬化症、小脳網膜血管芽腫(cerebelloretinal hemangioblastomatosis)、脳3叉神経血管症候群、中枢神経系性精神薄弱及び他の発生障害、脳性麻痺、神経骨格異常症、自律神経系障害、末梢神経疾患、皮膚筋炎及び多発性筋炎と、遺伝性、代謝性、内分泌性、及び中毒性ミオパシーと、重症筋無力症、周期性四肢麻痺と、気分性及び不安性の障害、分裂病性疾患を含む精神病と、季節性障害(SAD)と、静座不能、健忘症、緊張病、糖尿病性ニューロパシー、錐体外路性終末欠陥症候群、ジストニー、分裂病性精神障害、帯状疱疹後神経痛、及びトゥレット病と、進行性核上麻痺、皮質基部変性(corticobasal degeneration)及び家族性の前頭側頭性健忘症とが含まれ、心血管疾患の中には、動静脈瘤、アテローム硬化、高血圧、脈管炎、レイノー病、静脈奇形、動脈解離、静脈瘤、血栓静脈炎及び静脈血栓、血管の腫瘍や、血栓崩壊、バルーン血管形成術(balloon angioplasty)、血管置換術、および大動脈冠動脈バイパス術移植手術の合併症や、うっ血性心不全、虚血性心疾患、狭心症、心筋梗塞、高血圧性心疾患、変性弁膜性心疾患、石灰化大動脈弁狭窄症、先天性2尖大動脈弁、僧帽弁輪状石灰化(mitral annular calcification)、僧帽弁脱出、リウマチ熱、リウマチ性心疾患、感染性心内膜炎、非細菌性血栓性心内膜炎、全身性エリテマトーデスの心内膜炎、カルチノイド心疾患、心筋症、心筋炎、心膜炎、腫瘍性心疾患、先天性心臓疾患、及び心臓移植の合併症が含まれ、胃腸疾患の中には、嚥下障害、消化性食道炎、食道痙攣、食道狭窄、食道癌、消化不良、消化障害、胃炎、胃癌、食欲不振、悪心、嘔吐、胃不全麻痺、洞または幽門の浮腫、腹部アングナ、胸焼け、胃腸炎、イレウス、腸管感染、消化性潰瘍、胆石症、胆嚢炎、胆汁うっ滞、膵臓炎、膵臓癌、胆道疾患、肝炎、高ビリルビン血症、硬変症、肝臓の受動性うっ血、ヘパトーム、感染性大腸炎、潰瘍性大腸炎、潰瘍性直腸炎、クローン病、ホイップル病、マロリー ヴァイス症候群、結腸癌、結腸閉塞、過敏性腸症候群、短小腸症候群、下痢、便秘、胃腸出血、及び後天性免疫不全症候群(AIDS)腸症、黄疸、肝性脳症、肝腎症候群、肝炎、肝脂肪症、血色素症、ウィルソン病、1-アンチトリプシン欠損症、ライ症候群、原発性硬化性胆管炎、肝梗塞、門脈循環閉塞及び血栓、小葉中心壊死、肝臓紫斑病、肝静脈血栓、肝静脈閉塞症、子癇前症、子癇、妊娠性急性肝脂肪、妊娠性肝臓内胆汁うっ滞と、結節性再生及び腺腫、癌腫を含む肝癌とが含まれ、自己免疫/炎症性の疾患の中には、炎症及び日光性角化症、後天性免疫不全症候群(AIDS)及び副腎機能不全、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫性多腺性内分泌カンジダ性外胚葉ジストロフィ(APECED)、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、糖尿病、肺気腫、リンパ球毒素性一時性リンパ球減少症、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、心筋または心膜炎、骨関節炎、骨粗しょう症、肺炎、乾癬、ライター症候群、リウマチ様関節炎、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、原発性血小板血症、血小板減少症、潰瘍性大腸炎、ウェルナー症候群、癌合併症、血液透析、体外循環、ウイルス感染症、細菌感染症、真菌感染症、寄生虫感染症、原虫感染症、蠕虫感染症、外

10

20

30

40

50

傷が含まれ、代謝障害の中には、糖尿病、肥満症、および骨粗鬆症などが含まれ、感染症の中には、アデノウイルス及びアレナウイルス、ブニヤウイルス、カリチウイルス、コロナウイルス、フィロウイルス、ヘパドナウイルス、ヘルペスウイルス、フラビウイルス、オルソミクソウイルス、パルボウイルス、パポバウイルス、パラミキソウイルス、ピコルナウイルス、ポックスウイルス、レオウイルス、レトロウイルス、ラブドウイルス、トガウイルスに分類されるウイルス病原体による感染が含まれる。

【0154】

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G C R E C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、G C R E C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。

10

【0155】

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G C R E C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G C R E C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。

【0156】

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G C R E C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、G C R E C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。

【0157】

更なる実施例では、G C R E C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G C R E C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、上記した増殖異常、神経の疾患、心血管疾患、胃腸疾患、自己免疫/炎症性の疾患、代謝障害、およびウイルス感染が含まれる。一実施態様では、G C R E C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G C R E C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運搬機構として間接的に用いられ得る。

20

【0158】

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G C R E C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、G C R E C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。

30

【0159】

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、相補的な配列、ベクターを別の好適な治療薬と組み合わせることもできる。当業者は、従来の医薬原理にしたがって併用療法で用いる好適な治療薬を選択可能である。治療薬との組み合わせにより、上に列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。この方法を用いて少ない量の各薬剤で医薬効果をあげることが可能であり、広範囲な副作用の可能性を低減し得る。

【0160】

G C R E C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G C R E C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G C R E Cと特異的に結合するものを同定が可能である。G C R E C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フラグメント、及びF a b発現ライブラリによって作られたフラグメントが含まれる。但し、これらに限定されるものではない。治療用には、中和抗体（即ち、二量体の形成を阻害するもの）が特に好ましい。

40

【0161】

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、G C R E C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

50

免疫応答を高めることもできる。このようなアジュバントにはフロイントアジュバント、水酸化アルミニウムなどのミネラルゲルアジュバント、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、及びジニトロフェノールなどの界面活性剤が含まれるが、これらに限定されるものではない。ヒトに用いられるアジュバントの中では、BCG (*bacilli Calmette-Guérin*) 及び *Corynebacterium parvum* が特に好ましい。

【0162】

GCREC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或いはペプチド、またはそれらの断片は、天然のタンパク質のアミノ酸配列の一部と同一であることが望ましい。GCRECアミノ酸の短いストレッチは、KLH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。

10

【0163】

GCREC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。これらの技術には、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EBV-ハイブリドーマ技術が含まれるが、これらに限定されるものではない(例えば、Kohler, G. ら. (1975) *Nature* 256:495-497; Kozbor, D. ら. (1985) *J. Immunol. Methods* 81-8-42; Cote, R. J. ら. (1983) *Proc. Natl. Acad. Sci.* 80:2026-2030; Cole, S. P. ら. (1984) *Mol. Cell Biol.* 62:109-120を参照)。

20

【0164】

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、Morrison, S. L. 他. (1984) *Proc. Natl. Acad. Sci.* 81-4851-4855; Neuberger, M. S. 他. (1984) *Nature* 312:604-608; Takeda, S. ら. (1985) *Nature* 314:452, 454を参照)。別法では、当分野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、GCREC特異性一本鎖抗体を生成する。関連する特異性を備えるが別のイデオタイプの組成の抗体は、ランダムな組み合わせの免疫グロブリンライブラリから鎖混合によって生成することもできる(例えば、Burton D. R. (1991) *Proc. Natl. Acad. Sci.* 88:11120-3を参照)。

30

【0165】

抗体は、リンパ球集団の中の *in vivo* 産生を誘発することによって、また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示されているような、高度に特異的な結合試薬のパネルをスクリーニングすることによって、得ることもできる(例えば、Orlandi, R. 他. (1989) *Proc. Natl. Acad. Sci.* 86:3833-3837; Winter, G. 他. (1991) *Nature* 349:293-299を参照)。

40

【0166】

GCREC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、このような断片には、抗体分子のペプシン消化によって生成されるF(ab')に断片と、F(ab')に断片のジスルフィド架橋を減じることによって生成されるFab断片が含まれるが、これらに限定されるものではない。別法では、Fab発現ライブラリを作製することによって、所望の特異性とモノクローナルFab断片の迅速且つ容易な同定が可能となる(例えば、Huse, W. D. ら. (1989) *Science* 254:1275-1281を参照)。

50

【0167】

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。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、G C R E Cとその特異性抗体との間の複合体調整の計測が含まれる。二つの非干渉性G C R E C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、競合的な結合アッセイも利用することができる(P o u n d、前出)。

【0168】

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、G C R E C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G C R E C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G C R E C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医薬の K_a は、G C R E C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G C R E Cエピトープに単一特異的なモノクローナル抗体医薬の 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 / m o lの高親和性抗体医薬は、G C R E C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 / m o lの低親和性抗体医薬は、G C R E C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(immunopurification)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) Antibodies, Volume I: A Practical Approach. 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 . a n d C r y e r , A . (1 9 9 1) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY)。

【0169】

ある下流での適用におけるこのような医薬品の品質及び適性を調べるために、ポリクローナル抗体医薬の抗体価及び結合活性を更に評価する。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医薬は一般に、G C R E C抗体複合体を沈殿させなければならない処理に用いられる。様々な適用例における抗体の特異性及び抗体価、結合活性、抗体の品質や使用法の指針は一般に入手可能である。(例えば、C a t t y , 前出、及びC o l i g a n 他、前出を参照)。

【0170】

本発明の別の実施例では、G C R E C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、治療目的で使用することができる。ある実施態様では、G C R E C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(D N A及びR N A、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、センスまたはアンチセンスオリゴヌクレオチドまたは大きな断片が、G C R E C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。

【0171】

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できる(例えば、S l a t e r , J . E . 他(1998) J. Allergy Clin. Immunol. 102(3): 469-475; and S c a n l o n , K . J . 他(1995) 9(13): 1288-1296.を参照)。また、アンチセンス配列は、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(例えば、M i l l e r , A

. D. (1990) *Blood* 76:271; Ausubel, 前出; Ucker t, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347を参照)。その他の遺伝送達機構には、リポソーム系、人工的なウイルスエンベロープ、及び当分野で周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. 他 (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M. C. 他 (1997) *Nucleic Acids Res.* 25(14):2730-2736. を参照)。

【0172】

本発明の別の実施例では、GCREC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。遺伝子治療は、(i) 遺伝子欠損症(例えば、X染色体連鎖遺伝(Cavazzana-Calvo, M. 他 (2000) *Science* 288:669-672)によって特徴づけられる重度の複合型免疫欠損(SCID)-X1)、遺伝性アデノシン-デアミナーゼ(ADA)欠損症(Blaese, R. M. 他 (1995) *Science* 270:475-480; Bordignon, C. 他 (1995) *Science* 270:470-475)に関連する重度の複合型免疫欠損、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(thalassaemia)、家族性高コレステロール血症、第VIII因子若しくは第IX因子欠損による血友病(Crystal, R. G. (1995) *Science* 270:404-410; Verma, I. M. and Somia. N. (1997) *Nature* 389:239-242)を治療したり、(ii) 条件的致死性遺伝子産物(例えば、細胞増殖の制御不能による癌の場合)を発現させたり、及び(iii) 細胞内の寄生虫(例えば、ヒト免疫不全ウイルス(HIV)(Baltimore, D. (1988) *Nature* 335:395-396; Poeschbla, E. 他 (1996) *Proc. Natl. Acad. Sci. USA.* 93:11395-11399)や、B型若しくはC型肝炎ウイルス(HBV、HCV)、Candida albicans及びParacoccidioides brasiliensis等の真菌寄生虫、Plasmodium falciparum及びTrypanosoma cruzi等の原虫寄生体)に対する防御機能を有するタンパク質を発現させて行うことができる。GCREC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GCREC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。

【0173】

本発明の更なる実施例では、GCRECの欠損による疾患や異常症は、GCREC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GCREC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な導入技術には、(i) 個々の細胞内へのDNAのマイクロインジェクション、(ii) 金粒子の打ち込み、(iii) リポソーム仲介性トランスフェクション、(iv) 受容体仲介性遺伝子導入、及び(v) DNAトランスポソン(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. Recipon (1998) *Curr. Opin. Biotechnol.* 9:445-450)の使用が含まれる。

【0174】

GCREC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、PCDNA 3.1、EPITAG、PRCCMV2、PREP、PVAXベクター(Inv

itrogen, 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。GCREC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. and 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. and 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/ミフェプリストン誘導性プロモーター(Rossi, F.M.V. and H.M. Blau, 前出)、または(iii)正常な個体に由来するGCREC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。

10

【0175】

市販のリポソーム形質転換キット(例えば、Invitrogenが販売しているPERFECT LIPID及び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である。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. 他(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、これらの標準的な哺乳動物トランスフェクションプロトコルを変更する必要がある。

20

【0176】

本発明の別の実施例では、GCREC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(LTR)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GCRECをコードするポリヌクレオチドと、(ii)好適なRNAパッケージングシグナルと、(iii)追加のレトロウイルス・シス作用性RNA配列及び効率的なベクターの増殖に必要なコーディング配列を伴うRev応答性エレメント(RRE)とからなるレトロウイルスベクターを作製して治療することができる。レトロウイルスベクター(例えば、PFB及びPFBNEO)はStratagene社から入手可能であり、公表データ(Riviere, I. 他.(1995) Proc. Natl. Acad. Sci. U.S.A. 92:6733-6737)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。このベクターは、VSVg(Armentano, D. 他(1987) J. Virol. 61:1647-1650; Bender, M.A. 他(1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. 他(1998) J. Virol. 72:8463-8471; Zufferey, R. 他(1998) J. Virol. 72:9873-9880)等の乱交雑エンベロープタンパク質若しくは標的細胞上の受容体に対する親和性を有するエンベロープ遺伝子を発現する好適なベクター産生細胞系(VPCL)において増殖される。RIGGに付与された米国特許第5,910,434号(「Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant」)において、レトロウイルスパッケージング細胞系を得るための方法が開示

30

40

50

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。レトロウイルスベクターの増殖、ある細胞集団（例えば、CD4⁺T細胞）の形質導入、並びに形質導入した細胞を患者に戻す方法は、遺伝子治療の分野では周知であり、多数の文献に記載されている（Raniga, U. 他（1997）*J. Virol.* 71:7020-7029; Bauer, G. 他（1997）*Blood* 89:2259-2267; Bonyhadi, M.L.（1997）*J. Virol.* 71:4707-4716; Raniga, U. 他（1998）*Proc. Natl. Acad. Sci. U.S.A.* 95:1201-1206; Su, L.（1997）*Blood* 89:2283-2290）。

【0177】

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、GCRECの発現に関連する1或いは複数の遺伝子異常を有する細胞にGCREC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は当分野では周知である。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の中に導入するために可変性であることが証明された（Csete, M. E. 他（1995）*Transplantation* 27:263-268）。使用できる可能性のあるアデノウイルスベクターが、米国特許第5,707,618号（「Adenovirus vectors for gene therapy」）に記載されており、引用することを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また、Antinozzi, P.A. 他（1999）*Annu. Rev. Nutr.* 19:511-544; and Verma, I.M. and N. Somia（1997）*Nature* 18:389:239-242を参照し、引用することをもって本明細書の一部とする。

【0178】

別法では、ヘルペス系遺伝子治療の送達系を用いて、GCRECの発現に関連する1或いは複数の遺伝子異常を有する標的細胞にGCREC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（HSV）系のベクターは、HSV親和性の中枢神経細胞にGCREC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分野では周知である。複製適格性の単純疱疹ウイルス（HSV）I型系のベクターは、霊長類の眼にレポーター遺伝子を送達するために用いられてきた（Liu, X. 他（1999）*Exp. Eye Res.* 169:385-395）。HSV-1ウイルスベクターの作製は、DeLucaに付与された米国特許第5,804,413号（*Herpes simplex virus swains for gene transfer*）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。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コントロールの下で、細胞に導入される少なくとも1つの内在性遺伝子を含むゲノムからなる組換えHSV d92についての記載がある。また上記特許には、ICP4、ICP27及びICP22のために除去される組換えHSV株の作製及び使用方法が開示されている。HSVベクターについては、Goins, W.F. 他（1999）*J. Virol.* 73:519-532 and Xu, H. 他（1994）*Dev. Biol.* 163:152-161を参照し、引用することをもって本明細書の一部とする。クローニングされたヘルペスウイルス配列の操作や、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トランスフェクト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分野で周知の技術である。

【0179】

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GCREC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）の生物学的な研究が広範に行われ、遺伝子伝達ベクター（gene transfer vector）

10

20

30

40

50

がSFVゲノムに基づいていることが分かった(Garoff, H. and 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が完全長のゲノムRNAより高いレベルで複製されるため、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対してカプシドタンパク質が過剰に産生される。同様に、GCRECをコードする配列をウイルステ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GCRECをコードするRNAが産生され、高いレベルでGCRECが合成される。通常はウイルス感染は2~3日以内の細胞溶解に関係するが、シンドビスウイルス(SIN)の変異体を有するハムスターの正常な腎細胞(BHK-21)の持続的な感染を確立する能力は、ウイルスの溶解性の複製が遺伝子治療に適用できるように好適に変更することが可能であることを示唆している(Dryga, S. A. 他. (1997) *Virology* 228: 74 - 83)。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GCRECを導入することができる。ある集団における細胞のサブセットの特定の形質導入には、形質導入する前に細胞のソーティングを必要とする場合がある。ウイルスの感染性cDNAクローンの操作、ウイルスcDNA及びRNAのトランスフェクション、並びにウイルスの感染方法は当分野で周知である。

10

【0180】

例えば開始部位から約-10から約+10までの転写開始部位に由来するオリゴヌクレオチドを用いて、遺伝子の発現を阻害することが可能である。同様に、三重らせん塩基対合法を用いて阻害することができる。三重らせん構造は、二重らせんがポリメラーゼ、転写因子、または調節分子の結合のために十分に広がるのを阻止するため有用である。三重式DNAを用いる最近の治療の進歩は文献に記載されている(例えば、Gee, J. E. ら. (1994) *In: Huber, B. E. 及び B. I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY*を参照)。相補的な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の翻訳を阻止するように設計できる。

20

【0181】

酵素性RNA分子であるリボザイムは、RNAの特異的切断を触媒するために用いることができる。リボザイム作用の機構には、相補的な標的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が含まれ、ヌクレオチド鎖切断が続く。例えば、GCREC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。

30

【0182】

任意の潜在的RNA標的内の特異的なリボザイム切断部位が、後続の配列GUA、GUU、及びGUCを含むリボザイム切断部位に対して、標的分子をスキヤニングすることによって初めに同定される。一度同定されると、切断部位を含む標的遺伝子の領域に対応する15個から20個のリボヌクレオチドの短いRNA配列を、オリゴヌクレオチドの機能を不全にする二次的な構造の特徴について評価することが可能である。候補標的の適合性も、リボヌクレアーゼ保護アッセイを用いて、相補的なオリゴヌクレオチドとのハイブリダイゼーションの容易性をテストすることによって評価することが可能である。

40

【0183】

本発明の相補的なリボ核酸分子及びリボザイムは、当分野で周知の方法を用いて、核酸分子の合成のために作製することができる。これらの方法には、固相ホスホラミダイト化合物などのオリゴヌクレオチドを化学的に合成する方法が含まれる。別法では、RNA分子が*in vitro*及び*in vivo*でGCRECをコードするDNA配列の転写によって生成され得る、このようなDNA配列はT7またはSP6等の好適なRNAポリメラーゼプロモータを用いて、種々のベクターの中に組み入れることが可能である。別法では、

50

相補的なRNAを構成的または誘導的に合成するこれらのcDNA作製物は、細胞株、細胞、または組織の中に導入することができる。

【0184】

RNA分子を修飾することによって、細胞内の安定性を高め、半減期を長くすることができる。可能な修飾には、分子の5'及び/または3'端部でのフランキング配列の追加、または分子のバックボーン内のホスホジエステル結合の代わりにホスホロチオネートまたは2' Oメチルを用いる修飾が含まれるが、これらに限定されるものではない。PNAの生成に固有のこの概念は、内在性の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、及びウリジンのアセチル-、メチル-、チオ-、及び同様の修飾形態だけでなく、イノシン、キューエオシン(queosine)、及びワイプトシン(wybutosine)などの従来のものでない塩基を含めることによって、これらの分子の全体に拡大することができる。

10

【0185】

本発明の更なる実施例は、GCREC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。特定のポリヌクレオチドの発現の変化に有効な化合物には、限定するものではないが、特定のポリヌクレオチド配列と相互作用可能な非高分子化学物質、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、転写因子やその他のポリペプチド転写調節因子が含まれる。有効な化合物は、ポリヌクレオチド発現のインヒビター或いはエンハンサーとして作用し、ポリヌクレオチドの発現を変化させ得る。従って、GCRECの発現または活性の増加に

20

【0186】

特定のポリヌクレオチドの発現の変化の有効性を調べるために、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することができる。試験化合物は、有効な化合物の化学修飾を含む当分野で周知の任意の方法で得ることができる。このような方法は、ポリヌクレオチドの発現を変化させる場合、一般に市販されている或い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、標的ポリヌクレオチドの化学的及び/または構造的

特性に基づいて化合物を合理的にデザインする場合、更に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。GCREC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、*in vitro*細胞遊離系または再構成生化学系が含まれ得る。GCRECをコードするポリヌクレオチドの発現における変化は、当分野で周知の任意の方法でアッセイする。通常、GCREC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、特定の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の収量を定量し、その値が1或い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おける基準となり得る。試験化合物に曝露されるポリヌクレオチドの発現の変化が検出される場合は、ポリヌクレオチドの発現の変化に試験化合物が有効であることを示している。特定のポリヌクレオチドの発現の変化に有効な化合物を調べるために、例えば *Schizosaccharomyces pombe* 遺伝子発現系 (Atkins, D. 他 (1999) 米国特許第5,932,435号、Arndt, G.M. 他 (2000) *Nucleic Acids Res.* 28:E15) または HeLa 細胞等のヒト細胞株 (Clarke, M.L. 他 (2000) *Biochem. Biophys. Res. Commun.* 268:8-13) を用いてスクリーニングする。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を調べるための、各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、及び修飾オリゴヌクレオチド)の組み合わせライブラリのスクリーニ

30

40

50

ングを含む (Bruice, T.W. 他 (1997) 米国特許第5,686,242号、Bruice, T.W. 他 (2000) 米国特許第6,022,691号)。

【0187】

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でき、in vivo、in vitro、及びex vivoでの使用に等しく適している。ex vivoでの治療の場合、患者から採取された肝細胞の中にベクターを導入して、自家移植で同じ患者に戻すためにクローニング増殖される。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運搬は、当分野で周知の方法を用いて実行することができる (例えば、Goldman, C.K. 他 (1997) Nature Biotechnology 15:462-66:を参照)。

10

【0188】

上記したいかなる治療方法も、例えば、ヒト、イヌ、ネコ、ウシ、ウマ、ウサギ及びサルなどの哺乳動物を含む、治療が必要な全ての被験者に適用できる。

【0189】

本発明の別の実施例は、上記した全ての治療効果のために、医学上認められる担体と共に医薬品或いは無菌組成物の投与に関連する。このような組成物は、GCREC、GCRECの抗体、擬態、アゴニスト、アンタゴニスト、またはGCRECのインヒビターなどからなる。この組成物は、単体で、或いは安定剤などの1種類以上の別の薬剤と共に、無菌の生体適合性医薬品担体に投与することができる。このような医薬品担体には、生理食塩水、緩衝食塩水、ブドウ糖、及び水などが含まれるがこれらに限定されるものではない。この組成物は、単独或いは薬物またはホルモンなどの別の薬剤と共に投与することができる。

20

【0190】

本発明に用いられる組成物は、様々な経路を用いて投与するが可能である。この経路には、経口、静脈内、筋肉内、動脈内、骨髄内、クモ膜下、心室内、経皮、皮下、腹腔内、鼻腔内、腸内、局所、舌下、または直腸が含まれるがこれらに限定されるものではない。

【0191】

肺投与用の組成物は、液状または乾燥粉末状に調製することができる。このような組成物は通常、患者が吸入する直前にエアロゾル化する。小分子 (例えば、従来の低分子量有機薬剤) の場合には、速効製剤のエアロゾル輸送が当分野で周知である。高分子 (例えばより大きなペプチドやタンパク質) の場合には、肺の肺胞領域を介する肺輸送の技術が近年向上したため、インスリン等の薬剤を実際に血中に輸送することが可能となった (Patton, J.S. 他, 米国特許第5,997,848号等を参照)。肺輸送は、針注射を用いないで投与できるという点で優れており、潜在的に有毒な浸透エンハンサーが必要でなくなる。

30

【0192】

本発明に用いる好適な組成物には、目的を達成するため、効果的な量の活性処方成分を含む組成物が含まれる。当業者は、十分に自身の能力で効果的な服用量を決めることができる。

【0193】

GCREC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ボソーム製剤は、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、GCRECまたはその断片をHIV Tat-1タンパク質の陽イオンN末端部に結合することもできる。このようにして作製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されることが確認されている (Schwarze, S.R. 他 (1999) Science 285:1569-1572)。

40

【0194】

どのような組成物であっても、治療に効果的な薬用量は、初めは、例えば腫瘍細胞の腫瘍細胞アッセイで、或いは動物モデルのどちらかで推定することができる。通常、動物モデ

50

ルには、マウス、ウサギ、イヌ、サル、またはブタなどが用いられる。動物モデルはまた、好適な濃縮範囲及び投与の経路を決めるのに用いることができる。このような治療をもとに、ヒトへの有益な薬用量及び投与経路を決定することができる。

【0195】

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GCRECまたはその断片、GCRECの抗体、GCREC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、たとえば、ED₅₀（服用に対して集団の50%に医薬的効果がある用量）またはLD₅₀（服用に対して集団の50%に致命的である用量）統計を計算するなど、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決定することができる。毒性効果と治療効果との薬用量比は治療指数であり、LD₅₀/ED₅₀と示すことができる。高い治療指数を示す組成物が望ましい。細胞培養アッセイ及び動物実験から得られたデータが、ヒトへの適用のために、薬用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薬用量は、毒性を殆ど或いは全く含まず、ED₅₀を含む血中濃度の範囲であることが望ましい。薬用量は、用いられる投与形態及び患者の感受性、投与の経路によって、この範囲内で様々である。

10

【0196】

正確な薬用量は、治療が必要な患者に関する要素を考慮して、実務者によって決められるであろう。薬用量及び投与は、効果的なレベルの活性成分を与えるため或いは所望の効果を維持するために調節される。薬用量の要素として考慮されるものには、疾患の重症度、患者の一般的な健康状態、年齢、体重、及び患者の性別、投与の時間及び頻度、併用する薬剤、反応感受性、及び治療に対する応答が含まれる。作用期間が長い組成物は、三日か四日に一度、一週間に一度、二週間に一度、特定の製剤の半減期及びクリアランス率によって左右され、投与され得る。

20

【0197】

通常薬用量は投与の経路によって異なるが、約0.1~100,000µgまでの最大約1グラムまでである。特定の薬用量及び運搬の方法に関するガイダンスは文献に記載されており、一般に当分野の実務者は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とは異なったヌクレオチドの製剤を利用するであろう。同様に、ポリヌクレオチドまたはポリペプチドの運搬は、特定の細胞、状態、位置などに対して特異的であろう。

30

【0198】

（診断）

別の実施例では、GCRECに特異的に結合する抗体が、GCRECの発現によって特徴付けられる疾患の診断、またはGCRECやGCREC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。診断に有用な抗体は、治療のところで記載した方法と同じ方法で製剤される。GCREC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GCRECを検出する方法が含まれる。これらの抗体は、修飾をして或いはしないで使用され、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され得る。当分野で周知の種々のレポーター分子が用いられるが、その内の幾つかは上記した。

40

【0199】

GCRECを測定するためのELISA,RIA,及びFACS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GCREC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GCRECの発現の値は、複合体の形成に適した条件の下、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GCRECに対する抗体とを結合させることによって決定する。標準的な複合体形成の量は、測光法(photometric)などの種々の方法で定量され得る。被験者のGCREC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。基準値と被験者との間の偏差が、診断の指標となる。

50

【0200】

本発明の別の実施例によれば、GCREC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。用いられるポリヌクレオチドには、オリゴヌクレオチド配列、相補的なRNA及びDNA分子、及びPNA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GCREC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、GCRECの存在の有無、更に過剰な発現を調べ、治療中のGCREC値の調節を監視する。

【0201】

一実施形態では、GCREC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PCRプローブを用いたハイブリダイゼーションによって、GCREC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GCRECをコードする自然界の配列のみを同定するかどうか、或いはアレルや関連配列を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。

10

【0202】

プローブはまた、関連する配列の検出に利用され、GCREC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、DNAあるいはRNAが可能であり、SEQ ID NO: 22 - 42の配列、或いはGCREC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。

20

【0203】

GCRECをコードするDNA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、GCREC及びGCREC誘導体をコードするポリヌクレオチド配列をmRNA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。このようなベクターは市販されており、当業者には周知であり、好適なRNA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、in vitroでRNAプローブを合成するために用いられる。ハイブリダイゼーションプローブは、例えば³²P或いは³⁵Sなどの放射性核種、或いはアビジン/ビオチン(biotin)結合系によってプローブに結合されたアルカリホスファターゼなどの酵素標識等の種々のレポーターの集団によって標識され得る。

30

【0204】

GCRECをコードするポリヌクレオチド配列を用いて、GCREC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、このような疾患には増殖異常、神経の疾患、心血管疾患、胃腸疾患、自己免疫/炎症性の疾患、代謝障害、およびウイルス感染が含まれ、増殖異常の中には日光性角化症及びアテローム性動脈硬化、滑液包炎、硬変、肝炎、混合型結合組織病(MCTD)、骨髄線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症、並びに腺癌及び白血病、リンパ腫、黒色腫、骨髄腫、肉腫、及び奇形癌、具体的には、副腎、膀胱、骨、骨髄、脳、乳房、頸部、胆嚢、神経節、消化管、心臓、腎臓、肝臓、肺、筋肉、卵巣、脾臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、子宮の癌などが含まれ、神経の疾患の中には、癲癇、虚血性脳血管障害、脳卒中、大脳新生物、アルツハイマー病、ピック病、ハンチントン病、痴呆、パーキンソン病及びその他の錐体外路障害、筋萎縮性側索硬化及びその他の運動ニューロン障害、進行性神経性筋萎縮症、色素性網膜炎、遺伝性運動失調、多発性硬化症及び他の脱髄疾患、細菌性及びウイルス性髄膜炎、脳膿瘍、硬膜下蓄膿症、硬膜外膿瘍、化膿性頭蓋内血栓性静脈炎、脊髄炎及び神経根炎、ウイルス性中枢神経系疾患と、クールー及びクロイツフェルトヤコブ病、ゲルストマン症候群、Gerstmann-Strausler-Scheinker症候群を含むプリオン病と、致死性家族性不眠症、神経系性栄養病及び代謝病、神経線維腫症、結節硬化症、小脳網膜血管芽腫(cereb

40

50

e l l o r e t i n a l h e m a n g i o b l a s t o m a t o s i s)、脳3叉神経血管症候群、中枢神経系性精神薄弱及び他の発生障害、脳性麻痺、神経骨格異常症、自律神経系障害、末梢神経疾患、皮膚筋炎及び多発性筋炎と、遺伝性、代謝性、内分泌性、及び中毒性ミオパシーと、重症筋無力症、周期性四肢麻痺と、気分性及び不安性の障害、分裂病性疾患を含む精神病と、季節性障害(SAD)と、静座不能、健忘症、緊張病、糖尿病性ニューロパシー、錐体外路性終末欠陥症候群、ジストニー、分裂病性精神障害、帯状疱疹後神経痛、及びトゥレット病と、進行性核上麻痺、皮質基部変性(cortico basal degeneration)及び家族性の前頭側頭性健忘症とが含まれ、心血管疾患の中には、動静脈瘤、アテローム硬化、高血圧、脈管炎、レイノー病、静脈奇形、動脈解離、静脈瘤、血栓静脈炎及び静脈血栓、血管の腫瘍や、血栓崩壊、バルーン血管形成術(balloon angioplasty)、血管置換術、および大動脈冠動脈バイパス術移植手術の合併症や、うっ血性心不全、虚血性心疾患、狭心症、心筋梗塞、高血圧性心疾患、変性弁膜性心疾患、石灰化大動脈弁狭窄症、先天性2尖大動脈弁、僧帽弁輪状石灰化(mitral annular calcification)、僧帽弁脱出、リウマチ熱、リウマチ性心疾患、感染性心内膜炎、非細菌性血栓性心内膜炎、全身性エリテマトーデスの心内膜炎、カルチノイド心疾患、心筋症、心筋炎、心膜炎、腫瘍性心疾患、先天性心臓疾患、及び心臓移植の合併症が含まれ、胃腸疾患の中には、嚥下障害、消化性食道炎、食道痙攣、食道狭窄、食道癌、消化不良、消化障害、胃炎、胃癌、食欲不振、悪心、嘔吐、胃不全麻痺、洞または幽門の浮腫、腹部アングナ、胸焼け、胃腸炎、イレウス、腸管感染、消化性潰瘍、胆石症、胆嚢炎、胆汁うっ滞、膵臓炎、膵臓癌、胆道疾患、肝炎、高ビリルビン血症、硬変症、肝臓の受動性うっ血、ヘパトーム、感染性大腸炎、潰瘍性大腸炎、潰瘍性直腸炎、クローン病、ホイップル病、マロリー ヴァイス症候群、結腸癌、結腸閉塞、過敏性腸症候群、短小腸症候群、下痢、便秘、胃腸出血、及び後天性免疫不全症候群(AIDS)腸症、黄疸、肝性脳症、肝腎症候群、肝炎、肝脂肪症、血色素症、ウィルソン病、 - 1 - アンチトリプシン欠損症、ライ症候群、原発性硬化性胆管炎、肝梗塞、門脈循環閉塞及び血栓、小葉中心壊死、肝臓紫斑病、肝静脈血栓、肝静脈閉塞症、子癇前症、子癇、妊娠性急性肝脂肪、妊娠性肝臓内胆汁うっ滞と、結節性再生及び腺腫、癌腫を含む肝癌とが含まれ、自己免疫/炎症性の疾患の中には、炎症及び日光性角化症、後天性免疫不全症候群(AIDS)及び副腎機能不全、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫性多腺性内分泌カンジダ性外胚葉ジストロフィ(APECED)、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、糖尿病、肺気腫、リンパ球毒素性一時性リンパ球減少症、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、心筋または心膜炎、骨関節炎、骨粗しょう症、膵炎、乾癬、ライター症候群、リウマチ様関節炎、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、原発性血小板血症、血小板減少症、潰瘍性大腸炎、ウェルナー症候群、癌合併症、血液透析、体外循環、ウイルス感染症、細菌感染症、真菌感染症、寄生虫感染症、原虫感染症、蠕虫感染症、外傷が含まれ、代謝障害の中には、糖尿病、肥満症、および骨粗鬆症などが含まれ、感染症の中には、アデノウイルス及びアレナウイルス、ブニヤウイルス、カリチウイルス、コロナウイルス、フィロウイルス、ヘパドナウイルス、ヘルペスウイルス、フラビウイルス、オルソミクソウイルス、パルボウイルス、パポーバウイルス、パラミキソウイルス、ピコルナウイルス、ポックスウイルス、レオウイルス、レトロウイルス、ラブドウイルス、トガウイルスに分類されるウイルス病原体による感染が含まれる。GCRC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、PCR法、ディップスティック(dipstick)、ピン(pin)、ELISA式アッセイ、及び変異GCRC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。このような質的或いは量的方法は、当分野では周知である。

10

20

30

40

50

【0205】

ある実施態様では、G C R E C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。G C R E C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の下、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。好適な培養期間の後、サンプルを洗浄し、シグナルを定量して基準値と比較する。患者のサンプルのシグナルの量が、制御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G C R E C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を用いて、動物実験、臨床試験、或いは個人の患者の治療を監視における、特定の治療効果を推定することが可能である。

10

【0206】

G C R E C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の下、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、G C R E C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。標準的なハイブリダイゼーションは、正常な被験者から得た値と周知の量の実質的に精製されたポリヌクレオチドが用いられる実験からの値とを比較することによって定量可能である。正常なサンプルから得た標準的な値を、疾患の症状を示す被験者から得た値と比較可能である。基準値と被験者の値との偏差を用いて罹患しているかどうかを決定する。

20

【0207】

疾患の存在が確定され、治療プロトコルが開始されると、ハイブリダイゼーションアッセイを通常ペースで繰り返して、被験者における発現のレベルが正常な患者に示される値に近づき始めたかどうかを推定することが可能である。繰り返し行ったアッセイの結果を、数日から数ヶ月の期間の治療の効果を見るのに用いることができる。

【0208】

癌では、個体からの生体組織における異常な量の転写物が、疾患の発生の素因を示し、また実際に臨床的症状が出る前に疾患を検出する方法を提供することが可能である。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或いは積極的な治療法を早くから利用して、癌の発生または進行を防ぐことが可能となる。

30

【0209】

G C R E C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。このようなオリゴマーは、化学的な合成、酵素を用いた生成、或いは*in vitro*で生成され得る。オリゴマーは、好ましくはG C R E Cをコードするポリヌクレオチドの断片、或いはG C R E C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、最適な条件の下、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な条件の下、近縁のD N A或いはR N A配列の検出及び/または定量のため用いることが可能である。

【0210】

或る実施態様において、G C R E C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、ヒトの先天性または後天性遺伝病の原因となる場合が多いヌクレオチドの置換、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、一本鎖立体構造多型(S S C P)及び蛍光S S C P(f S S C P)法が含まれる。S S C Pでは、G C R E C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。このD N Aは、例えば病変或いは正常な組織、生検サンプル、体液等に由来し得る。このD N A内のS N Pは、一本鎖形状のP C R産物の2次及び3次構造に差異を生じさせる。この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標識することによって、D N Aシーケンシング装置などのハイスループット機器でアンプリマー(a m

40

50

plimer)の検出をすることが可能になる。更に、インシリコSNP (in silico SNP: isSNP)と呼ばれる配列データベース分析法は、共通のコンセンサス配列の構築に用いられる個々の重複するDNA断片の配列を比較することによって、多型を同定することができる。これらのコンピュータベースの方法は、DNA配列クロマトグラムの自動分析及び統計モデルを用いたシーケンシングエラーや研究室でのDNAの調整に起因する配列のばらつきを排除する。別法では、例えばハイスループットのMASSARRAYシステム (Sequenom, Inc., San Diego CA)を用いた質量分析によりSNPを検出し、特徴付ける。

【0211】

GCRECの発現を定量するために用いられ得る方法には、ヌクレオチドの放射標識或いはビオチン標識、調節核酸の相互増幅 (coamplification)、及び標準的な曲線に結果が加えられたものが含まれる (例えば、Melby, P. C.ら (1993) J. Immunol. Methods, 159:235-44; Duplaa, C.ら (1993) Anal. Biochem. 229-236を参照)。多数のサンプルの定量速度は、ハイスループット型のアッセイを用いることで速くなるであろう。このアッセイでは、目的のオリゴマーやポリヌクレオチドが様々な希釈液中に含まれ、分光光度法或いは非色応答によって定量が迅速である。

【0212】

更に別の実施例では、本明細書に記載した任意のポリヌクレオチド配列に由来するオリゴヌクレオチドまたはより長い断片を、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。マイクロアレイを、上記したように多数の遺伝子の相対的な発現レベルを同時にモニタリングする転写イメージング技術に用いることができる。マイクロアレイはまた、遺伝子変異、突然変異及び多型の同定に用いることができる。この情報を用いて、遺伝子機能を決定し、疾患の遺伝的根拠を解明し、疾患を診断し、遺伝子発現に関連する疾病の進行/後退をモニタリングし、疾患の治療における治療薬の開発や活性のモニタリングを行うことができる。特に、患者にとって最適かつ有効な治療法を選択するために、この情報を用いて患者の薬理ゲノムプロファイルを作成することができる。例えば、患者の薬理ゲノムプロファイルに基づいて、患者に対して極めて効果的でありながら副作用を殆ど示さない治療薬を選択することができる。

【0213】

別の実施例では、GCREC、GCRECの断片、GCREC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、上記のようにタンパク質間相互作用、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リング及び測定することが可能である。

【0214】

特定の実施例は、或る組織または細胞型の転写イメージを生成する本発明のポリヌクレオチドの使用に関連する。転写イメージは、特定の組織または細胞型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、所定の条件下で所定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される (Seilliamer 他、米国特許第5,840,484号の "Comparative Gene Transcript Analysis" を参照。この特許に言及することを以って本明細書の一部とする)。従って、特定の組織または細胞型の転写物または逆転写物の全てに本発明のポリヌクレオチドまたはその相補配列をハイブリダイズすることにより、転写イメージが生成され得る。或る実施例では、本発明のポリヌクレオチドまたはその相補配列がマイクロアレイ上に複数のエレメントのサブセットを構成するハイスループット型でハイブリダイゼーションさせる。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となり得る。

【0215】

転写イメージは、組織、細胞株、生検サンプル、またはその他の生体サンプルから単離した転写物を用いて生成し得る。従って、転写イメージは、組織または生検サンプルの場合

には *in vivo*、または細胞株の場合には *in vitro* における遺伝子発現を反映する。

【0216】

本発明のポリヌクレオチドの発現プロファイルを示す転写イメージはまた、合成化合物または天然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され得る。全ての化合物は、作用及び毒性の機構を示唆する、頻繁に分子フィンガープリント若しくは毒性シグネチャ (signature) と称されるような特徴的な遺伝子発現パターンを引き起こす (Nuwaysir, E. F. 他 (1999) *Mol. Carcinog.* 24:153-159、Steiner, S. and N. L. Anderson (2000) *Toxicol. Lett.* 112-113:467-471、また言及することを以って本明細書の一部とする)。試験化合物が、毒性を有する既知の化合物のシグネチャと同一のシグネチャを有する場合には、毒性特性を共有している可能性が高い。フィンガープリントまたはシグネチャが、より多く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れば、より有用かつ正確になる。理想としては、発現のゲノム全域にわたって測定し、最高品質のシグネチャを提供することである。任意の試験化合物によっても発現が変化しない遺伝子も同様に重要である。それは、これらの遺伝子の発現レベルを用いて残りの発現データを標準化することができるためである。標準化処理は、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。毒性シグネチャのエLEMENTへの遺伝子機能を割り当てることは毒性機構の解明に役立つが、毒性の予測につながるシグネチャの統計的な一致には遺伝子機能の知識は必要ではない (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.htm> で入手可能である)。従って、毒性シグネチャを用いる毒性スクリーニングにおいて、全ての発現した遺伝子配列を含めることは重要でありまた望ましいことである。

10

20

【0217】

一実施例では、試験化合物の毒性は、核酸を含有する生体サンプルをその試験化合物で処理して評価する。処理した生体サンプル中で発現した核酸は、本発明のポリヌクレオチドに特異的な1若しくは複数のプローブでハイブリダイズさせ、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することができる。処理した生体サンプル中の転写レベルを、非処理生体サンプル中のレベルと比較する。両サンプルの転写レベルの差が、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唆する。

30

【0218】

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型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」という用語は、或る特定の組織または細胞型におけるタンパク質発現の包括的パターンを指す。プロテオームを構成する各タンパク質は、個々に更なる分析をすることができる。プロテオーム発現パターン即ちプロファイルは、所定の条件下で所定の時間に発現したタンパク質の数及びそれらの相対的な存在量を定量することにより分析する。従って、ある細胞のプロテオームのプロファイルは、特定の組織または細胞型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、このような分離は2次元ゲル電気泳動によって行う。この2次元ゲル電気泳動法では、まず、1次元の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、次に、2次元の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従って分離する (前出のSteiner and Anderson)。これらのタンパク質は、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染色剤を用いてゲルを染色して、分散した個別の位置にあるスポット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済みまたは未処理のいずれかの生体サンプルから得られる等位置にあ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

40

50

を調べる。スポット内のタンパク質は、例えば化学的または酵素的に切断した後、質量分析す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、好適には少なくとも5個の連続するアミノ酸残基であるその部分的な配列を、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。

【0219】

プロテオームのプロファイルは、G C R E C に特異的な抗体を用いて G C R E C 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、マイクロアレイ上の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露して各アレイエレメントへのタンパク質結合レベルを検出することによりタンパク質発現レベルを定量する (Lueking, A. ら. (1999) Anal. Biochem. 270:103-111, Mendozze, L. G. ら. (1999) Biotechniques 27:778-788)。検出は当分野で既知の様々な方法で行うことができ、例えば、チオール反応性またはアミノ反応性蛍光化合物を用いてサンプル中のタンパク質を反応させ、各アレイエレメントにおける蛍光結合の量を検出し得る。

10

【0220】

プロテオームレベルでの毒性シグネチャも中毒学的スクリーニングに有用であり、転写レベルでの毒性シグネチャと並行して分析するべきである。或る組織における或るタンパク質では、転写物の存在量とタンパク質の存在量との相関性が低いことがあるため (Anderson, N. L. and J. Seilhamer (1997) Electrophoresis 18:533-537)、プロテオーム毒性シグネチャは、転写イメージにはそれ程影響しないがプロテオームのプロファイルを変化させる化合物の分析において有用たり得る。更に、体液中での転写の分析は、mRNA が急速に分解するため困難である。したがって、このような場合にはプロテオームのプロファイル作成はより信頼でき、情報価値がある。

20

【0221】

別の実施例では、試験化合物の毒性は、タンパク質を含む生体サンプルをその試験化合物で処理して評価する。処理された生体サンプル中で発現したタンパク質を分離して、各タンパク質の量が定量できるようにする。各タンパク質の量を、未処理生体サンプル中の対応するタンパク質の量と比較する。両サンプル中のタンパク質の量の差は、処理されたサンプル中の試験化合物に対する毒性反応を示唆する。個々のタンパク質は、それらのアミノ酸残基をシーケンシングし、これらの部分配列を本発明のポリペプチドと比較することで同定する。

30

【0222】

別の実施例では、試験化合物の毒性は、タンパク質を含む生体サンプルをその試験化合物で処理することにより評価する。生体サンプルから得たタンパク質を、本発明のポリペプチドに特異的な抗体と共にインキュベートする。その抗体により認識されたタンパク質の量を定量する。処理された生体サンプル中のタンパク質の量を、未処理生体サンプル中のタンパク質の量と比較する。両サンプルのタンパク質量の差が、処理サンプル中の試験化合物に対する毒性反応を示唆する。

40

【0223】

当分野で周知の方法でマイクロアレイを準備して使用し、分析する。(例えば、Brennan, T. M. 他 (1995) 米国特許第5,474,796号; Schena, M. 他 (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler 他 (1995) PCT出願番号WO95/251116; Shalon, D. 他 (1995) PCT出願番号WO95/35505; Heller, R. A. 他 (1997) Proc. Natl. Acad. Sci. 94:2150-2155; 及び Heller, M. J. 他 (1997) 米国特許第5,605,662号を参照)。様々なタイプのマイクロアレイが周知であり、詳細については、DNA Microarrays: A Practical A

50

pproach, M. Schena, ed. (1999) Oxford University Press, Londonに記載されている。また、この文献を引用することを以って本明細書の一部とする。

【0224】

本発明の別の実施例ではまた、GCRC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。コーディング配列または非コーディング配列の何れかを用いることができるが、或る例では、コーディング配列より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間にコーディング配列が保存されていることにより、染色体マッピング時に望ましくない交差ハイブリダイゼーションが生じる可能性がある。この配列は、特定の染色体、染色体の特定領域または人工の染色体、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.ら(1997) Nat Genet. 15:345-355、Price, C. M. (1993) Blood Rev. 7:127-134、Trask, B. J. (1991) Trends Genet. 7:149-154等を参照)。一度マッピングすると、本発明の核酸配列を用いて、例えば病状の遺伝と特定の染色体領域やまたは制限断片長多型(RFLP)の遺伝とが相関するような遺伝子連鎖地図を作成可能である(Lander, E. S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357を参照)。

【0225】

in situ蛍光ハイブリダイゼーション(FISH)は、他の物理的及び遺伝子地図データと相関し得る(例えば、Heinz-Ulrich, 他による(1995) in Meyers, 前出, pp. 965-968を参照)。遺伝子地図データの例は、種々の科学誌あるいはOnline Mendelian Inheritance in Man(OMIM)のワールドワイドウェブのサイトで見つけることができる。物理的な染色体地図上のGCRC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。

【0226】

染色体標本のin situハイブリダイゼーション、及び確定した染色体マーカーを用いた結合分析などの物理的マッピング技術を用いて、遺伝子地図を拡張することもできる。マウスなどの別の哺乳動物の染色体上に遺伝子を配置させることにより、たとえ正確なヒト染色体の位置が分かっていなくても、関連するマーカーが明らかになる場合が多い。この情報は、位置クローニング或いは別の遺伝子発見技術を用いて遺伝的疾患の研究をしている研究者にとって価値がある。疾患や症候群に關与する1つ或いは複数の遺伝子の位置が、例えば血管拡張性失調症の11q22-23などの特定の遺伝子領域に遺伝子結合によって大まかに決定されると、その領域に対するどの配列マッピングも、さらなる調査のための関連する遺伝子或いは調節遺伝子を表す(例えば、Gatti, R. A. 他による(1988) Nature 336:577-580を参照)。また、目的の本発明のヌクレオチド配列を用いて、正常者、保有者、即ち感染者の間の、転位置、反転などによる染色体位置の違いを検出することもある。

【0227】

本発明の別の実施例では、GCRC、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。このようなスクリーニングに用いる断片は、溶液に遊離、固体支持物に固定、細胞の表面上に保持、或いは細胞内に存在する。GCRC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。

【0228】

薬剤スクリーニングに用いる別の方法は、目的のタンパク質に対して、好適な結合親和性

10

20

30

40

50

を有する化合物のスクリーニング処理能力を高めるために用いられる（例えば、Geysen, 他による（1984）PCT出願番号WO84/03564を参照）。この方法では、相当な数の異なる小さな試験用化合物が、プラスチックピン或いは他の基板の上に合成される。試験用化合物は、GCREC、或いはその断片と反応してから洗浄される。次に、結合されたGCRECが、当分野で周知の方法で検出される。精製されたGCRECはまた、前記した薬剤をスクリーニングする技術に用いられるプレート上で直接被覆することもできる。別法では、非中和抗体を用いて、ペプチドを捕らえ、固体支持物に固定することもできる。

【0229】

別の実施例では、GCRECと結合可能な中和抗体がGCREC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、GCREC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。

10

【0230】

別の実施例では、発展途上の分子生物学技術にGCREC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。

【0231】

当分野の技術者であれば、更なる説明がなくても前述の説明だけで最大限に本発明を利用できるであろう。したがって、以下に記載する実施例は、例示目的であって本発明を限定

20

【0232】

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60/186,854号、60/188,384号、60/190,453号、および同第60/190,730号に言及することをもって本明細書の一部とする。

【0233】

（実施例）

1 cDNAライブラリの作製

インサイトcDNAはLIFESEQ GOLD データベース（Incyte Genomics, Palo Alto CA）に含まれているcDNAライブラリに由来し、表4の列5に示されている。組織の一部を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する一方、この組織の別の一部をホモジナイズしてフェノールに溶解するか、或いはTRIZOL（Life Technologies）、グアニジニウムイソチオシアネート及びフェノールの単相溶液などの好適な変性剤の混合液に溶解した。この溶解物を塩化セシウムにおいて遠心分離によって、或いはクロロホルムで抽出した。イソプロパノール或いは酢酸ナトリウムのどちらかとエタノール、或いは別の方法でこの溶解物からRNAを沈殿させた。

30

【0234】

RNAの純度を高めるために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、DNA分解酵素でRNAを処理する。殆どのライブラリでは、オリゴd(T)連結常磁性粒子（Promega）またはOLIGOTEXラテックス粒子（QIAGEN, Valencia CA）、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、POLY(A)PURE mRNA精製キット（Ambion, Austin TX）などの別のRNA単離キットを用いて組織溶解物から直接単離した。

40

【0235】

ある場合には、Stratagene社にRNAを提供し、Stratagene社が対応するcDNAライブラリを作製した。そうでない場合は、UNIZAPベクターシステム（Stratagene）またはSUPERSCRIPT プラスミドシステム（Life Technologies）を用いて当分野で周知の推奨方法または類似の方法で

50

cDNAを合成してcDNAライブラリを作製した。(例えば、Ausubel, 1997, 前出, ユニット5.1-6.6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に結合させてから、好適な1つの制限酵素或いは複数の制限酵素でcDNAを消化した。殆どのライブラリでは、SEPHACRYL S 1000またはSEPHAROSE CL2B、SEPHAROSE CL4Bカラムクロマトグラフィー(Amersham Pharmacia Biotech)、アガロースゲル電気泳動法によってcDNAの大きさ(300~1000bp)を選択した。PBLUESCRIPTプラスミド(Stratagene)またはpSPORT1プラスミド(Life Technologies)、pcDNA2.1プラスミド(Invitrogen Carlsbad CA)、PBK-CMVプラスミド(Stratagene)、pINCYプラスミド(Incyte Pharmaceuticals, Palo Alto CA)、またはそれらの誘導体などの好適なプラスミドのポリリンカーの適合性制限酵素部位にcDNAを結合させた。この組換えプラスミドを、Stratagene社のXL1-Blue, XL1-Blue MRF、SOLR、またはLife Technologies社のDH5 またはDH10B、ELECTROMAX DH10Bを含むコンピテント大腸菌細胞に導入し組み込んだ。

【0236】

2 cDNAクローンの単離

上記実施例1に記載したように得たプラスミドを、UNIZAPベクターシステム(Stratagene)或いは細胞溶解を利用した*in vivo*切除によって宿主細胞から回収した。MagicまたはWIZARD Minipreps DNA精製システム(Promega)、及びAGTC Miniprep精製キット(Edge Biosystems, Gaithersburg MD)、QIAGEN社のQIAWELL 8 Plasmid、QIAWELL 8 Plus Plasmid、QIAWELL 8 Ultra Plasmid 精製システム、REAL Prep 96プラスミドキットの内の少なくとも1つを用いてプラスミドを精製した。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。

【0237】

別法では、ハイスループットの直接結合PCR法によっ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を単一反応混合液で行った。サンプルを処理してから384-ウェルプレートに移して保管し、増幅したプラスミドDNAの濃度をPICOGREEN色素(Molecular Probes, Eugene OR)及びFluoroskan II蛍光スキャナ(Labsystems Oy, Helsinki, Finland)を用いて蛍光定量的に測定した。

【0238】

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インサイトcDNAを、以下に示すようにシークエンシングした。cDNAのシークエンシング反応は標準的な方法で行うか、またはHYDRAマイクロディスペンサー(Robbins Scientific)或いはMICROLAB 2200(Hamilton)液体移送装置と共にABICATALYST 800(PE Biosystems)サーマルサイクラー或いはPTC-200 thermal cycler(MJ Research)などのハイスループット装置を用いて行った。cDNAのシークエンシング反応は、Amersham Pharmacia Biotech社の試薬、またはABI PRISM BIGDYE Terminator cycle sequencing ready reactionキット(PE Biosystems)などのABIシークエンシングキットに含まれる試薬を用いて行った。cDNAシークエンシングの反応物の電気泳動的による分離及び標識したポリヌクレオチドの検出は、MEGABACE 1000 DNAシークエンシングシス

テム (M o l e c u l a r D y n a m i c s)、標準 A B I プロトコル及び塩基対呼び出しソフトウェアを用いる A B I P R I S M 3 7 3 または 3 7 7 シークエンシングシステム (P E B i o s y s t e m s)、または当分野で周知のその他の配列解析システムを用いて行った。c D N A 配列内の読み枠は、標準的な方法 (A u s u b e l , 1 9 9 7 , 前出 , u n i t 7 . 7) を用いて決定した。c D N A 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した。

【 0 2 3 9 】

インサイト c D N A に由来する本ポリヌクレオチド配列の確認は、B L A S T、動的プログラミング、およびジヌクレオチドの分布による解析 (d i n u c l e o t i d e n e a r e s t n e i g h b o r a n a l y s i s) に基づいたプログラム及びアルゴリズムを用いて、ベクター、リンカー、およびポリ A 配列を取り除き、更にあいまいな塩基対をマスクすることで行った。次に、インサイト c D N A 配列およびそれらの翻訳を、公共のデータベースである G e n B a n k の霊長類、げっ歯類、哺乳類、脊椎動物、および真核生物のデータベース、および B L O C K S、P R I N T S、D O M O、P R O D O M、および P F A M などの隠れマルコフモデル (H M M) を基にしたタンパク質ファミリーのデータベースから選択した配列に対して問合せた (H M M は、遺伝子ファミリーのコンセンサス主構造を分析する確率的手法である。例えば、E d d y , S . R . (1 9 9 6) C u r r . O p i n . S t r u c t . B i o l . 6 : 3 6 1 - 3 6 5 を参照)。このような問合せは、B L A S T、F A S T A、B L I M P S、および H M M R に基づいたプログラムを用いて行った。インサイト c D N A 配列を組み立てて、完全長ポリヌクレオチド配列を作製した。或いは、G e n B a n k c D N A s、G e n B a n k E S T、ステッチ配列 (s t i t c h e d s e q u e n c e)、ストレッチ配列 (s t r e t c h e d s e q u e n c e s)、または G e n s c a n - 推定コード配列 (実施例 4 および 5 を参照) を用いて、インサイト c D N A 群を完全長の配列に伸長した。配列の組み立ては、P h r e d、P h r a p、および C o n s e d に基づいたプログラムを用いて行い、G e n e M a r k、B L A S T、および F A S T A に基づいたプログラムを用いて、オープンリーディングフレームを決定するべく c D N A 群をスクリーニングした。完全長のポリヌクレオチド配列を翻訳して対応する完全長ポリペプチド配列を得た。別法では、本発明のポリヌクレオチドは、完全長翻訳ポリヌクレオチドの任意のメチオニン残基から始まり得る。次に、完全長ポリペプチド配列を G e n B a n k タンパク質データベース (g e n p e p t)、S w i s s P r o t、B L O C K S、P R I N T S、D O M O、P R O D O M、P r o s i t e、および P F A M などの隠れマルコフモデル (H M M) に基づいたタンパク質ファミリーデータベースに対して問合せて分析した。これらの完全長ポリヌクレオチド配列はまた、M A C D N A S I S P R O ソフトウェア (H i t a c h i S o f t w a r e E n g i n e e r i n g , S o u t h S a n F r a n c i s c o C A) および L A S E R G E N E ソフトウェア (D N A S T A R) を用いて分析した。ポリヌクレオチド配列およびポリペプチド配列のアラインメントを、アラインメントした配列間のパーセント同一性も計算する M E G A L I G N マルチシークエン্সアラインメントプログラム (D N A S T A R) に組み込まれた C L U S T A L アルゴリズムによって指定されたデフォルトパラメータを用いて作成した。

【 0 2 4 0 】

表 7 は、インサイト c D N A および完全長配列の組み立て、および組み立てた配列の分析に利用したツール、プログラム、およびアルゴリズム、並びにそれらの説明、引用文献、閾値パラメータを簡単に示す。表 7 の列 1 は用いたツール、プログラム、およびアルゴリズム、列 2 はそれらの簡単な説明、列 3 は引用することで本明細書の一部とした引用文献、列 4 の記載されている部分は 2 つの配列の一致の程度を評価するために用いたスコア、確率値、およびその他のパラメータを示す (スコアが高くなれば高くなるほど即ち確率値が低ければ低いほど、配列間の相同性が高くなる)。

【 0 2 4 1 】

完全長のポリヌクレオチド配列およびポリペプチド配列の組み立て及び分析に用いる上記

のプログラムは、SEQ ID NO: 22 - 42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。

【0242】

4 ゲノムDNAからのコード配列の同定および編集

推定上のGタンパク質共役受容体は、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、様々な生物に由来するゲノムDNA配列を分析するための汎用遺伝子同定プログラムである（Burge, C. および S. Karlin (1997) J. Mol. Biol. 268 : 78 - 94、Burge, C. および S. Karlin (1998) Curr. Opin. Struct. Biol. 8 : 346 - 354を参照）。このプログラムは推定エキソンを連結して、メチオニンから停止コドンまで伸長した組み立てcDNA配列を構築する。Genscanにより得られる配列は、FASTAデータベースのポリヌクレオチド配列およびポリペプチド配列になる。Genscanによって一回で解析できる配列の最大長さは30kbに設定されている。これらのGenscan推定cDNA配列の内、どの配列がGタンパク質共役受容体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Gタンパク質共役受容体について問合せて分析した。潜在的なGタンパク質共役受容体が、Gタンパク質共役受容体としてアノテーションが付けられたインサイトcDNA配列に対する相同性を基に同定された。次に、これらの選択されたGenscan推定配列を、BLAST解析を用いてgeneptおよびgbpri公共データベースの配列と比較した。必要に応じて、Genscan推定cDNA配列を、geneptにおいてBLASTで最もヒットした配列と比較して、Genscan推定配列における余分なエキソンや省いてしまったエキソンなどのエラーを修正し、編集した。BLAST解析を用いてGenscan推定cDNA配列を含むインサイトcDNAまたは公共のcDNAを見つけ出すことにより、転写の証拠が得られる。インサイトcDNAがGenscan推定cDNA配列を含む場合、この情報を用いてGenscan推定配列を修正或いは確認できる。完全長ポリヌクレオチド配列は、実施例3に説明した組み立て方法でGenscan推定コード配列とインサイトcDNAおよび/または公共のcDNA配列を組み立てて作製した。別法では、完全長ポリヌクレオチド配列は、その全てが編集した或いは未編集のGenscan推定コード配列から作製した。

【0243】

5 cDNA配列データを用いたゲノム配列データの組み立て

ステッチ配列 (Stitched Sequence)

部分的なcDNA配列を、実施例4に記載したGenscan遺伝子同定プログラムによって推定されたエキソンで伸長した。実施例3に記載されたように組み立てられた部分的なcDNAをゲノムDNAにマッピングし、1或いは複数のゲノム配列から推定されるGenscanエキソンおよび関連cDNAを含む複数のクラスターに分けた。各クラスターを、グラフ理論および動的計画法に基づいたアルゴリズムを用いて、cDNAおよびゲノム情報を統合して分析し、可能性のあるスプライスバリエーションを生成して、続いて確認、編集、または伸長して完全長の配列を作製した。区間の全長が或るクラスターの2つ以上の配列に存在する配列区間を同定し、このように同定した区間は移行によるもので同等と考える。例えば、或る区間がcDNAおよび2つのゲノム配列のそれぞれに存在する場合、これら3つ全ての区間を同等と考える。この方法によって、関連性がないが連続するゲノム配列をcDNA配列によって繋ぎ合わせる。このようにして同定された区間を、親配列 (parent sequence) に沿って現われるようにステッチアルゴリズムを用いて順に縫い合わせ、可能な最長の配列および変異配列を作製する。或るタイプの親配列 (cDNAとcDNA、またはゲノム配列とゲノム配列) に沿って連結される区間と区間との繋ぎ合わせは、親配列のタイプが異なる (cDNAとゲノム配列) 連結より好ましい。得られたステッチ配列を翻訳し、BLAST解析でgeneptおよびgbpri

i 公共データベースにおける配列と比較した。GenScanによって推定された不適当なエキソンを、geneptにおいてBLASTで最もヒットした配列と比較して修正する。このような配列を更なるcDNA配列で伸長し、必要に応じてゲノムDNAで検査した。

【0244】

ストレッチ配列 (Stretched Sequence)

部分的なDNA配列をBLAST解析に基づいたアルゴリズムで完全長に伸長した。まず、実施例3に記載したように組み立てた部分的なcDNAを、BLASTプログラムを用いてGenBankの霊長類、げっ歯類、哺乳類、脊椎動物、および真核生物のデータベースなどの公共のデータベースに対して問い合わせた。次に、GenBankの相同性の最も高いタンパク質を、実施例4に記載したインサイトcDNA配列或いはGenScanエキソン推定配列の何れかとBLAST分析により比較した。得られた複数の高スコアのセグメント対(HSP)を用いてキメラタンパク質を作製し、GenBankの相同タンパク質上に翻訳した配列をマッピングした。元のGenBankタンパク質相同体に対して、キメラタンパク質に挿入や欠失が起こり得る。公共のヒトゲノムデータベースから相同ゲノム配列を探し出すために、GenBankタンパク質相同体、キメラタンパク質、またはそれらの両方をプローブとして用いた。このようにして、部分的なDNA配列を相同ゲノム配列の付加によりストレッチすなわち伸長した。完全な遺伝子を含んでいるか否かを判定するために得られたストレッチ配列を検査した。

10

【0245】

6 GCRECをコードするポリヌクレオチドの染色体マッピング

SEQ ID NO: 22-42を組み立てるために用いた配列を、BLAST及びSmith-Watermanアルゴリズムを用いて、インサイトLIFESEQ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。SEQ ID NO: 22-42と一致するこれらのデータベースの配列を、Phrap(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。Stanford Human Genome Center (SHGC)、Whitehead Institute for Genome Research (WIGR)及びGenethonなどの公共の情報源から入手できる放射線ハイブリッド(radiation hybrid)及び遺伝子マッピングのデータを用いて、クラスター化した配列がすでにマッピングされているかを調べる。クラスターにマッピングされた配列が含まれている場合は、そのクラスターの全ての配列(特定のSEQ ID NOを含む)をそのマッピング位置に割り当てた。

20

30

【0246】

遺伝子地図の位置は、範囲、区間、またはヒト染色体によって表される。センチモルガンで示したマッピング位置の範囲は、染色体の短腕(p)の末端から測定した(センチモルガン(cM)は、同一染色体上の遺伝子間の乗換え率に基づいた距離を表す単位である。平均すると、1cMはヒトの染色体の1メガベースに概ね等しいが、組換え率の高い部分と低い部分があるため、大きく変化し得る)。距離cMは、配列がそれぞれのクラスターに含まれている放射線ハイブリッドマーカーの境界を検出できるGenethonによってマッピングされた遺伝子マーカーに基づいている。NCBI「GeneMap99」(<http://www.ncbi.nlm.nih.gov/genemap>)などの公衆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。

40

【0247】

このようにSEQ ID NO: 23は、57.8~71.4センチモルガンの間隔でクロモソーム16へとマッピングされた。

【0248】

7 ポリヌクレオチド発現の分析

ノーザン分析は、遺伝子の転写物の存在を検出するために用いられる実験用技術であり、

50

特定の細胞種或いは組織からのRNAが結合されている膜への標識されたヌクレオチド配列のハイブリダイゼーションを伴う(例えば、Sambrook, 前出, 7章; 及び Ausubel, F.M., 他、前出, 4章及び16章を参照)。

【0249】

BLASTに用いる類似のコンピュータ技術を用いて、GenBank或いはLIFESEQ (Incyte Pharmaceuticals)のようなcDNAデータベース内の同一或いは関連する分子を検索する。この分析は多くの膜系ハイブリダイゼーションより非常に速度が速い。さらにコンピュータ検索の感度を変更して、任意の特定の一致が、厳密な一致或いは相同的一致の何れかとして分類されるかを確定することができる。検索の基準は、

10

【0250】

【数1】

(BLASTスコア×配列一致率)

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

として定義される積スコアである。積スコアは、0~100の標準化された値であり、以下のように求める。BLASTスコアにヌクレオチド配列の一致率を乗じ、その積を2つの配列の短い方の長さの5倍で除する。高スコアのセグメントの対(HSP)において一致する各塩基に+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、BLASTアラインメントの断片的重複と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%一致する場合にのみ得られる。積スコア70は、100%の同一性で重畳が70%であるか、或いは88%の同一性で重畳が100%であるかのいずれかの場合である。積スコア50は、100%の同一性で重畳が50%であるか、或いは79%の同一性で重畳が100%であるかの何れかの場合である。

20

【0251】

或いは、GCREC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、ある完全長の配列は、少なくとも部分的にインサイトcDNA配列をオーバーラップさせて組み立てられる(実施例3を参照)。各cDNA配列は、ヒト組織から作製されたcDNAライブラリに由来する。各ヒト組織は、心血管系、結合組織、消化系、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、血液および免疫系、肺、筋骨格系、神経系、膵臓、呼吸器系、感覚器官、皮膚、顎口腔系、分類不能/混合、または尿管などの1つの生物/組織のカテゴリーに分類される。各カテゴリーにおけるライブラリの数をカウントし、その合計数を全カテゴリーのライブラリ数で除す。同様に、各ヒト組織は、癌、細胞系、発生、炎症、神経、外傷、心血管、プール(pool)などの1つの疾患/症状のカテゴリーに分類され、各カテゴリーにおけるライブラリの数をカウントし、その合計数を全カテゴリーのライブラリ数で除す。得られるパーセンテージは、GCRECをコードするcDNAの疾患特異的な発現を反映する。cDNA配列およびcDNAライブラリ/組織の情報は、LIFESEQ GOLD データベース(Incyte Genomics, Palo Alto CA)から得ることができる。

30

40

【0252】

8 GCRECをコードするポリヌクレオチドの伸長

完全長のポリヌクレオチド配列は、完全長分子の好適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その完全長分子の好適な断片を伸長して作製した。一方のプライマーは既知の断片の5'の伸長を開始するために合成し、他方のプライマーは既知の断片の3'の伸長を開始するために合成した。開始プライマーは、OLIGO 4.06ソフトウェア(National Biosciences)或いは他の適切なプログラムを

50

用いて、約22個から約30個のヌクレオチドの長さで約50%以上のGC含量を有し、かつ約68~72の温度で標的配列にアニールするように設計した。ヘアピン構造及びプライマー-プライマー二量体が生じないようにヌクレオチドを伸長した。

【0253】

選択されたヒトcDNAライブラリを用いてこの配列を伸長した。2段階以上の伸長が必要な場合、若しくは望ましい場合は、追加或いはネスト化プライマーの組を設計する。

【0254】

当分野で既知の方法を利用したPCR法で高い忠実度で増幅した。PCRはPTC-200 thermal cycler (MJ Research, Inc.)用いて96ウェルブロックプレートで行った。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β-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 (Amersham Pharmacia Biotech)、ELONGASE酵素 (Life Technologies)、Pfu DNAポリメラーゼ (Stratagene)を含む。プライマーの組、PCI AとPCI 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秒

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 で保管

別法では、プライマーの組、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秒

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 で保管。

【0255】

各ウェルのDNA濃度は、1X TE及び0.5 µlの希釈していないPCR産物に溶解した100 µlのPICO GREEN定量試薬(0.25% (v/v) PICO GREEN; Molecular Probes, Eugene OR)を不透明な蛍光光度計プレート(Coming Costar, Acton MA)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。このプレートをFluoroskan II (Labsystems Oy, Helsinki, Finland)でスキャンして、サンプルの蛍光を計測してDNAの濃度を定量化する。反応混合物の5~10 µlのアリコットを1%のアガロースミニゲル上での電気泳動によって解析し、何れの反応物が配列を伸長することに成功したかを決定する。

【0256】

伸長したヌクレオチドを脱塩及び濃縮してから384ウェルプレートに移し、CviJIコレラウィルスエンドヌクラーゼ (Molecular Biology Research, Madison WI)で消化し、pUC18ベクター (Amersham Pharmacia Biotech)に再連結する前に音波処理またはせん断を行った。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に分離して断片を切断し、寒天をAgar ACE (Promega)で消化した。T4リガーゼ (New England Biolabs, Beverly MA)を用いて伸長したクローンをpUC18ベクター (Amersham Pharmacia Biotech)に再連結し、Pfu DNAポリメラーゼ (Stra

10

20

30

40

50

t a g e n e) で制限部位の伸び出しを処理して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LB / 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。

【0257】

細胞を溶解して、Taq DNAポリメラーゼ (Amersham Pharmacia Biotech) 及びPfu DNAポリメラーゼ (Stratagene) 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秒
- 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29回繰り返す
- ステップ6 72 で5分間
- ステップ7 4 で保管。

10

上記したようにPICO GREEN試薬 (Molecular Probes) でDNAを定量化した。DNA回収率の悪いサンプルは、上記した条件で再び増幅した。サンプルを20%のジメチルサルホサイド (dimethylsulphoxide) (1:2, v/v) で希釈し、DYENAMIC DIRECTキット (Amersham Pharmacia Biotech) またはABI PRISM BIGDYE Terminator cycle sequencing ready reactionキット (Applied Biosystems) を用いてシーケンシングした。

20

【0258】

同様に上述の手順で、完全長のポリヌクレオチド配列を検査したり、或いは完全長のポリヌクレオチド配列を利用して、この伸長のために設計したオリゴヌクレオチドと好適なゲノムライブラリを用いて5 調節配列を得た。

【0259】

9 個々のハイブリダイゼーションプローブの標識化及び使用法

SEQ ID NO: 22 - 42 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、cDNA、mRNA、またはゲノムDNAをスクリーニングする。約20塩基対からなるオリゴヌクレオチドの標識について特に記すが、より大きなcDNAフラグメントの場合でも基本的に同じ手順を用いる。オリゴヌクレオチドを、OLIGO4.06ソフトウェア (National Bioscience) のような最新式のソフトウェアを用いてデザインし、50 pmolの各オリゴマーと、250 μ Ciの[³²P]アデノシン三リン酸 (Amersham, Chicago, IL) 及びT4ポリヌクレオチドキナーゼ (DuPont NEN, Boston MA) とを組み合わせて用いることにより標識する。標識されたオリゴヌクレオチドを、SEPHADEX G-25超精細排除デキストランビードカラム (Amersham Pharmacia Biotech) を用いて実質的に精製する。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、次のエンドヌクレアーゼ、Ase I、Bgl II、Eco RI、Pst I、Xba I 或いはPvu II (DuPont NEN) の1つを用いて切断したヒトゲノム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用いる。

30

40

【0260】

各切断物からのDNAを、0.7%アガロースゲル上で分画して、ナイロン製メンブラン (Nytran Plus, Schleicher & Schuell, Durham NH) に転写する。ハイブリダイゼーションは40 で16時間かけて行う。非特異的シグナルを取り除くため、例えば、最大0.1x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の条件の下、プロットを順次室温にて洗浄する。ハイブリダイゼーションパターンをオートラジオグラフィー或いは別のイメージ化手段で視覚化して比較する。

【0261】

50

10 マイクロアレイ

マイクロアレイ上のアレイエレメントの連結または合成は、フォトリソグラフィ、ピエゾプリント（インクジェットプリンター、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板は、均一な非多孔性の固体とするべきである（Schena（1999）、前出）。推奨する基板には、シリコン、シリカ、スライドガラス、ガラスチップ及びシリコンウエハがある。別法では、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、熱や紫外線、または化学的或いは機械的な結合手段で基板の表面にエレメントを配置して結合させることができる。通常のアレイは利用可能な方法や機械を用いて作製でき、任意の適正な数のエレメントを含めることができる（Schena, M. 他（1995）Science 270: 467-470、Shalon, D. 他（1996）Genome Res. 6: 639-645、Marshall, A. and J. Hodgson（1998）Nat. Biotechnol. 16: 27-31.を参照）。

【0262】

完全長cDNA、発現遺伝子配列断片（EST）、或いはそれらの断片やオリゴマーが、マイクロアレイのエレメントとなり得る。ハイブリダイゼーションに好適な断片やオリゴマーを、LASERGENEソフトウェア（DNASTAR）などの当分野で周知のソフトウェアを用いて選択することが可能である。このアレイエレメントを、生体サンプル中のポリヌクレオチドとハイブリダイズさせる。生体サンプル中のポリヌクレオチドは、検出を容易にするために蛍光標識またはその他の分子タグに結合する。ハイブリダイゼーションの後、生体サンプルからハイブリダイズしなかった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けるハイブリダイゼーションを検出する。別法では、レーザー脱離及び質量スペクトロメトリー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程度及び相対的存在量は、算定することができる。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。

【0263】

組織または細胞サンプルの調製

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サンプルは、MMLV逆転写酵素、0.05 pg/μlのオリゴ（dT）プライマー（21mer）、1×第1鎖緩衝液、0.03単位/μlのRNアーゼインヒビター、500 μM dATP、500 μM dGTP、500 μM dTTP、40 μM dCTP、40 μM dCTP-Cy3（BDS）またはdCTP-Cy5（Amersham Pharmacia Biotech）を用いて逆転写する。この逆転写反応は、GEMBRIGHT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。特異的なコントロールポリ（A）⁺ RNAは、in vitro転写により非コーディング酵母ゲノムDNAから合成する。370 で2時間インキュベートした後、各反応サンプル（一方はCy3標識、他方はCy5標識）は、2.5 mlの0.5 M 水酸化ナトリウムで処理し、850 で20分間インキュベートし、反応を停止させてRNAを変性する。サンプルは、2つの連続するCHROMA SPIN 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つの反応サンプルを、1 mlのグリコーゲン（1 mg/ml）、60 mlの酢酸ナトリウム及び300 mlの100%エタノールを用いてエタノール沈殿させる。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 μl 5×SSC/0.2% SDS中で再懸濁する。

【0264】

マイクロアレイの準備

10

20

30

40

50

本発明の配列を用いて、アレイエレメントを作製する。各アレイエレメントは、クローン化 cDNA 挿入断片を含むベクターを含有する細菌性細胞から増幅する。PCR 増幅は、cDNA 挿入断片に隣接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 によって、1 ~ 2 ng の初期量から 5 µg を超える最終量までアレイエレメントを増幅する。増幅されたアレイエレメントは、SEPHACRYL - 400 (Amersham Pharmacia Biotech) を用いて精製する。

【0265】

精製したアレイエレメントを、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での洗浄と、0.1% の SDS 及びアセトン中で超音波による洗浄を行う。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% エタノール中の 0.05% アミノプロピルシラン (Sigma) でコーティングする。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。

10

【0266】

米国特許第 5,807,522 号に記載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。この特許に引用することを以って本明細書の一部とする。平均濃度が 100 ng / µl のアレイエレメント DNA 1 µl を高速機械装置により開放型キャピラリープリンティングエレメント (open capillary printing element) に充填する。次にこの装置が、スライド毎に約 5 nl のアレイエレメントサンプルを分注する。

20

【0267】

マイクロアレイには、STRATALINKER UV クロスリンカー (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% SDS で 1 回洗浄し、蒸留水で 3 回洗浄する。非特異的な結合部位は、リン酸緩衝生理食塩水 (PBS) (Tropix, Inc., Bedford MA) における 0.2% カゼイン中で 60 °C で 30 分間マイクロアレイをインキュベートし、その後上述したように 0.2% SDS 及び蒸留水で洗浄することによってブロックする。

【0268】

ハイブリダイゼーション

ハイブリダイゼーション反応液は、5 × SSC、0.2% SDS ハイブリダイゼーション緩衝液に Cy3 及び Cy5 標識した cDNA 合成産物を各 0.2 µg 含む 9 µl のサンプル混合体を含めたものである。サンプル混合液を、65 °C で 5 分間加熱し、マイクロアレイ表面上に一定量分注してから 1.8 cm² のカバーガラスで覆う。このアレイを、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ャンバーに移す。チャンバーの角に 140 µl の 5 × SSC を加えて、チャンバー内を湿度 100% に保持する。このアレイを含むチャンバーを、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 × SSC, 0.1% SDS) において 45 °C で 10 分間、第 2 洗浄緩衝液中 (0.1 × SSC) において 45 °C で 10 分間それぞれ 3 回洗浄し、その後乾燥させる。

30

40

【0269】

検出

レポーター標識されたハイブリダイゼーション複合体は、Cy3 を励起するための 488 nm、及び Cy5 を励起するための 632 nm のスペクトル線を生成し得る Innova 70 混合ガス 10 W レーザー (Coherent, Inc., Santa Clara CA) を備えた顕微鏡で検出する。20 倍の顕微鏡対物レンズ (Nikon, Inc., Melville NY) を用いて、アレイ上に励起レーザー光を集中させる。この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、対物レンズを通してラスタスキャンする。本実施例で用いた 1.8 cm × 1.8 cm のアレイは、20

50

μmの解像度でスキャンする。

【0270】

2つの異なるスキャンにおいて、混合ガスマルチラインレーザーは2つの蛍光体を連続的に励起する。放射された光は、波長に基づいて2つの蛍光体に対応する2つの光電子増倍管検出器(PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ)に分割される。アレイと光電子増倍管との間に配設された好適なフィルタを用いて信号をフィルタリングする。用いる蛍光体の最大発光は、Cy3では565nm、Cy5では650nmである。装置は両方の蛍光体からのスペクトルを同時に記録できるが、レーザー源に好適なフィルタを用いて、蛍光体1つにつき1回スキャンし、各アレイを通常2回スキャンする。

10

【0271】

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コントロール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配列を含め、その位置におけるシグナルの強度がハイブリダイズする種の重量比1:100,000に相関するようにする。異なる試料(例えば検査細胞及びコントロール細胞を代表する)からの2つのサンプルを、各々異なる蛍光体で標識し、他と異なって発現する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させる場合には、校正は2つの蛍光体を有する校正するcDNAのサンプルを標識して、ハイブリダイゼーション混合液に各々等量を加えて行う。

【0272】

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、リニア20色変換を用いてシグナル強度が青色(低シグナル)から赤色(高シグナル)までの擬似カラー範囲にマッピングされるイメージとして表示される。データはまた、定量的に分析される。2つの異なる蛍光体を同時に励起して測定する場合には、各蛍光体の発光スペクトルを用いて、先ずデータは蛍光体間の光学的漏話(重複発光スペクトルに起因する)に対して補正される。

20

【0273】

グリッドを蛍光シグナルイメージ上に重畳して、各スポットからのシグナルがグリッドの各エレメントに中央に位置するようにする。各エレメント内の蛍光シグナルを統合し、シグナルの平均強度に対応する数値を得る。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。

30

【0274】

1.1 相補的ポリヌクレオチド

GCREC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GCREC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~約30個の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、より小さな或いはより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GCREC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'配列から相補的な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GCRECをコードする転写物に結合するのを阻害する。

40

【0275】

1.2 GCRECの発現

GCREC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できる。細菌でGCRECが発現するために、抗生物質耐性及びcDNA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cDNAをサブクローニングする。このようなプロモーターには、lacオペレーター調節エレメントに関連するT5また

50

はT7バクテリオファージプロモーター及びtrp-lac(tac)ハイブリッドプロモーターが含まれるが、これらに限定されるものではない。組換えベクターを、BL21(DE3)などの好適な細菌宿主に形質転換する。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(IPTG)で誘発されるとGCRCを発現する。真核細胞でのGCRC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Autographica californica核多面性ウイルス(ACMNPV)を感染させて行う。バキュロウイルスの非必須ポリヘドリン遺伝子を、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、GCRCをコードするcDNA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cDNA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Spodoptera frugiperda(Sf9)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。(例えば、Engelhard, E. K. 他(1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. 他(1996) Hum. Gene Ther. 7:1937-1945.を参照)。

10

【0276】

殆どの発現系では、GCRCが、例えばグルタチオンSトランスフェラーゼ(GST)、またはFLAGや6-Hisなどのペプチドエピトープ標識で合成された融合タンパク質となるため、未精製の細胞溶解物からの組換え融合タンパク質の親和性ベースの精製が素早く1回で行うことができる。Schistosoma japonicumからの26キロダルトンの酵素GSTによって、タンパク質の活性及び抗原性を維持した状態で固定されたグルタチオンで融合タンパク質の精製が可能となる(Amersham Pharmacia Biotech)。精製の後、GST部分を特定の操作部位でGCRCからタンパク質的に切断できる。アミノ酸8個のペプチドであるFLAGで、市販のモノクローナル及びポリクローナル抗FLAG抗体(Eastman Kodak)を用いた免疫親和性の精製が可能となる。6個の連続するヒスチジン残基のストレッチである6-Hisによって、金属キレート樹脂(QIAGEN)で精製が可能となる。タンパク質の発現及び精製の方法は、Ausubel(1995, 前出, ch 10, 16)に記載されている。これらの方法で精製したGCRCを直接用いて以下の実施例の16、17、および18のアッセイを行うことができる。

20

30

【0277】

1.3 機能のアッセイ

GCRCの機能は、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GCRCをコードする配列の発現によって評価する。cDNAを、cDNA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サブクローニングする。このようなベクターには、pCMV SPORTTM(Life Technologies)及びpCR3.1(Invitrogen, Carlsbad, CA)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含んでいる。5~10µgの組換えベクターを、例えば内皮由来か造血由来のヒト細胞株にリポソーム製剤或いは電気穿孔法によって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1~2µg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、形質移入された細胞と形質移入されていない細胞とを区別できる。また、標識タンパク質の発現によって、cDNAの組換えベクターからの発現を正確に予想できる。このような標識タンパク質には、緑色蛍光タンパク質(GFP; Clontech)、及びCD64またはCD64-GFP融合タンパク質が含まれる。レーザー光学に基づいた技術を利用した自動流動細胞計測法(FCM)を用いて、GFPまたはCD64-GFPを発現する形質移入された細胞を同定し、その細胞のアポトーシス状態や他の細胞特性を評価する。また、FCMで、先行した或いは同時の細胞死の現象を診断する蛍光分子の取り込みを検出して計量する。これらの現象には、プロピジウムヨウ化物でのDNAの染色によって計測される核DNA内容物

40

50

の変化と、プロモデオキシウリジンの取り込み量の低下によって計測されるDNA合成の下方調節と、特異的な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のタンパク質の発現の変化と、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含まれる。流動細胞計測法は、Ormerod, M. G. による(1994) Flow Cytometry Oxford, New York, NY. に記載されている。

【0278】

遺伝子発現におけるGCRECの影響は、GCREC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)の保存された領域と結合する。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。GCREC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。

10

【0279】

1.4 GCREC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. 1816-3088-495を参照)または他の精製技術で実質的に精製されたGCREC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。

20

【0280】

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

【0281】

通常、約15残基の長さのオリゴペプチドを、Applied BiosystemsのABI 431Aペプチドシンセサイザー(PE Biosystems)を用いてfmoc法のケミストリにより合成し、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)を用いた反応によりKLH(Sigma-Aldrich, St. Louis MO)に結合させて、免疫原性を高める(例えば、前出のAusubel, 1995を参照)。フロイントの完全アジュバントにおいてオリゴペプチド-KLH複合体を用いてウサギを免疫化する。得られた抗血清の抗ペプチド活性及び抗GCREC活性を検査するには、ペプチドまたはGCREC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。

30

【0282】

1.5 特異的な抗体を用いる天然GCRECの精製

天然GCREC或いは組換えGCRECを、GCREC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GCREC抗体とを共有結合させることにより形成する。結合の後、そのレジンを製造者の使用説明書に従ってブロッキング処理し洗浄する。

40

【0283】

GCRECを含む培養液をイムノアフィニティークラムに通し、GCREC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

50

のカラムを洗浄する。そのカラムを、抗体とGCRECとの結合を切るような条件で（例えば、pH 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、GCRECを回収する。

【0284】

16 GCRECと相互作用する分子の同定

GCRECと相互作用する分子には、アゴニスト、アンタゴニスト、およびGタンパク質などのシグナル伝達物質に関与する分子が含まれる。GCRECまたはその断片を、¹²⁵I Bolton-Hunter 試薬（例えば、Bolton A. E. 及びW. M. Hunter (1973) Biochem. J. 133: 529を参照）で標識する。GCRECの断片には、例えば、3つの細胞外ループの内の1或いは複数の細胞がいループ、細胞外のN末端領域、または第3の細胞外ループを含む断片が含まれる。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GCRECと共にインキュベートし、洗浄して、標識したGCREC複合体を有する全てのウェルをアッセイする。様々なGCREC濃度で得られたデータを用いて、候補分子と結合したGCRECの数量及び親和性、会合についての値を計算する。

10

【0285】

別法では、GCRECと相互作用する分子を、Fields, S. 及びO. Song (1989, Nature 340: 245-246)に記載の酵母2-ハイブリッドシステム(yeast two-hybrid system)やMATCHMAKERシステム(Clontech)などの2-ハイブリッドシステムに基づいた市販のキットを用いて分析する。GCRECはまた、ハイスループット型の酵母2ハイブリッドシステムを使用するPATHCALLINGプロセス(CuraGen Corp., New Haven CT)に用いて、遺伝子の2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(Nandabalan, K. 他 (2000) 米国特許第6,057,101号)。

20

【0286】

実施例17および18に示すアッセイを用いて、潜在的なGCRECのアゴニストまたはアンタゴニストを、GCREC受容体活性の活性化または阻害について検査することができる。候補となる分子は、既知のGPCRのアゴニストまたはアンタゴニスト、ペプチドライブラリ、または組み合わせキメラライブラリから選択され得る。

30

【0287】

GCRECと、Gタンパク質などの細胞内シグナル伝達物質との相互作用を検出する方法は、オーファンGタンパク質結合7回膜貫通型受容体の内側セグメントすなわち細胞質ドメインが、既知のGタンパク質結合7回膜貫通型受容体の類似のドメインと置換するであろうという前提に基づいており、それを用いてそのオーファン受容体ドメインによって活性化されるGタンパク質および下流のシグナル伝達経路を同定する(Kobilka, B. K. ら (1988) Science 240: 1310-1316)。同様に、オーファン受容体のドメインを、融合タンパク質の一部としてクローニングし、それを用いて特定のGタンパク質との相互作用を実証するための結合アッセイを行う。研究結果より、Gタンパク質結合7回膜貫通型受容体の第3の細胞内ループがGタンパク質の相互作用およびシグナル伝達に重要であることが分かった(Conklin, B. R. ら (1993) Cell 73: 631-641)。例えば、GCRECの第3の細胞内ループに対応するDNA断片をPCR法で増幅して、pGEX(Pharmacia Biotech)などの融合ベクターにサブクローニングすることができる。この作製物を好適な細菌宿主に形質転換し、発現させ、その融合タンパク質を、グルタチオンセファロース4B(Pharmacia Biotec)アフィニティークロマトグラフィーを用いて細胞溶解物から精製する。

40

【0288】

in vitro結合アッセイのために、Gタンパク質を含む細胞抽出物を、50 mM Tris、pH 7.8、1 mM EGTA、5 mM MgCl₂、20 mM CHAP

50

S、20% グリセロール、10 μ gのアプロチニン、10 μ gのロイペプチン、及び20 μ lの50 mMフェニルメチルスルホニルフッ化物で抽出して準備する。この溶解物を、氷上で掻き混ぜながら45分間インキュベートし、23,000 \times gで15分間4で遠心分離し、上清を収集する。750 μ gの細胞抽出物を、グルタチオンSトランスフェラーゼ(GST)融合タンパク質ビーズと共に4で2時間インキュベートする。GSTビーズをリン酸緩衝生理食塩水で5回洗浄する。結合したGサブユニットを、百日咳毒素またはコレラ毒素での $[^3\text{H}]$ ADP-リボシル化によって検出する。SDSサンプルバッファー(4.6% (w/v) SDS、10% (v/v) β -メルカプトエタノール、20% (w/v) グリセロール、95.2 mM Tris-HCl、pH 6.8、0.01% (w/v) ブロムフェノールブルー)を追加して反応を停止させる。 $[^3\text{H}]$ ADP標識したタンパク質を10% SDS-PAGEゲル上で分離させ、オートラジオグラフィ法を実施する。これらの分離したタンパク質をニトロセルロース紙に移し、室温で1時間、blotting(5%脱脂粉末乳、50 mM Tris-HCl (pH 8.0)、2 mM CaCl₂、80 mM NaCl、0.02% NaN₃、および0.2% Nonidet P-40)で遮断し、次にGサブタイプ選択抗体(1:500; Calbiochem-Novabiochem)で1.5時間インキュベートする。3回洗浄した後、プロットをカラシペルオキシダーゼ(HRP)接合ヤギ抗ウサギ免疫グロブリン(1:2000, Cappel, Westchester PA)と共にインキュベートし、化学発光を利用したECL法(Amersham Corp.)によって視覚化する。

10

20

30

40

50

【0289】

1.7 GCRECの活性の実証

GCRECの活性についてのアッセイでは、細胞表面のGCRECの発現を測定する。GCRECをコードするcDNAを好適な哺乳動物細胞系に形質転換する。細胞表面タンパク質をビオチン標識する(e la Fuente, M. A. ら(1997) Blood 90: 2398-2405を参照)。GCREC特異的抗体を用いて免疫沈降を行い、免疫沈降したサンプルをドデシル硫酸ナトリウムポリアクリルアミドゲル電気泳動(SDS-PAGE)および免疫ブロッティング技術を用いて分析する。標識した免疫沈降物と標識していない免疫沈降物との割合が、細胞表面で発現したGCRECの量に比例する。

【0290】

別法では、GCRECの活性のアッセイは、リガンド/受容体仲介性の細胞増殖の変化を調べるプロトタイプアッセイに基づいている。このアッセイでは、Swissマウス3T3細胞におけるDNAの合成の速度を測定する。GCRECをコードするポリヌクレオチドを含むプラスミドを、当分野で周知のトランスフェクション法で静止状態の3T3培養細胞に加える。一過性にトランスフェクトされた細胞を、放射性DNA前駆体分子である $[^3\text{H}]$ チミジンの存在下でインキュベートする。次に、様々な量のGCRECリガンドを培養細胞に加える。ラジオアイソトープカウンタを用いて、酸沈降性DNAに組み込まれた $[^3\text{H}]$ チミジンの量を適当な時間の間測定する。組み込まれた量が新規に合成されたDNAの量に正比例する。少なくとも100倍のGCRECリガンドの濃度範囲に対して線形の線量効果曲線が、受容体の活性を表す。1 ml当たりの活性1単位を、50%の反応レベルが生じるGCRECの濃度と定義する。この場合、100%は、 $[^3\text{H}]$ チミジンが酸沈降性DNAに最大に組み込まれたことを表す(McKay, I. および I. Leigh, eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York NY, p. 73)。

【0291】

更なる別法では、GCRECの活性のアッセイは、GPCRファミリータンパク質がGタンパク質仲介性セカンドメッセンジャーシグナル伝達経路を変える能力に基づいている(例えば、cAMP; Gaudin, P. ら(1998) J. Biol. Chem

273 : 4990 - 4996)。完全長GCRECをコードするプラスミドを、当分野で周知の方法を用いて哺乳動物細胞系に形質転換する(例えば、チャイニーズハムスター卵巣細胞系(CHO)またはヒト胚腎細胞株(HEK-293))。形質転換細胞を12ウェルトレイの培養液で48時間増殖させ、次に培養液を捨て、付着している細胞をPBSで軽く洗浄する。次に、リガンドを含む培養液または含まない培養液で細胞を30分間インキュベートしてから培養液を捨て、1M過塩素酸で処理して細胞を溶解する。溶解産物中のcAMPのレベルを、当分野で周知の方法を用いてラジオイムノアッセイで測定する。リガンドに曝露しなかった細胞の溶解産物中におけるcAMPのレベルに対するリガンドに曝露した細胞の溶解産物中におけるcAMPのレベルの変化が、トランスフェクト細胞に存在するGCRECの量に比例する。

10

【0292】

イノシトールリン酸のレベルの変化を測定するために、1ウェル当たり 1×10^5 細胞を含む24ウェルプレートで細胞を増殖させ、イノシトールを含まない培地および $[^3\text{H}]$ ミオイノシトール、 $2 \mu\text{Ci}$ /ウェルでインキュベートする。培地を除去してから、細胞を10mM LiClを含むバッファーで洗浄し、次にリガンドを加える。この反応は、過塩素酸を加えて停止させる。イノシトールリン酸をDowex AG1-X8 (Bio-Rad) 陰イオン交換樹脂上で抽出および分離し、標識したイノシトールリン酸の全てを液体シンチレーションでカウントする。リガンドに曝露しなかった細胞からの標識したイノシトールリン酸に対するリガンドに曝露した細胞からの標識したイノシトールリン酸のレベルの変化が、トランスフェクト細胞に存在するGCRECの量に比例する。

20

【0293】

1.8 GCRECリガンドの同定

CHO(チャイニーズハムスター卵巣)またはHEK293(ヒト胚腎)などの真核生物細胞系でGCRECを発現させる。これらの細胞系は、GPCRの発現に好適であり、発現したGCRECと下流のエフェクターとの機能的な結合を可能にする種々のGタンパク質を含む。候補リガンドの存在下での発現された受容体の活性化を調べるために形質転換細胞をアッセイする。活性の測定は、サイクリックAMPや Ca^{2+} などの細胞内セカンドメッセンジャーの変化を調べて行う。測定には、当分野で周知の標準的な方法、或いはレポーター遺伝子アッセイを用いて行うことができる。このレポーター遺伝子アッセイでは、発光タンパク質(例えば、ホタルルシフェラーゼや緑色蛍光タンパク質)が、活性化された受容体によるプロテインキナーゼCの刺激に応答するプロモーターの転写制御下にある(Milligan, G.ら(1996) Trends Pharmacol. Sci. 17 : 235 - 237)。アッセイ技術は、これらのセカンドメッセンジャー系の両方に利用でき、FLIPR蛍光定量的プレート読出し装置(Molecular Devices)を用いた、アデニルシクラーゼ活性化FlashPlateアッセイ(NEN Life Sciences Products)またはFluo-4 AM (Molecular Probes)などの蛍光 Ca^{2+} インジケータなどのマルチウェルプレート型におけるハイスループット読出しを可能にする。生理学的に関連するセカンドメッセンジャー経路がわかっていない場合、ホスホリパーゼCおよび Ca^{2+} の動員に関する経路をGCRECのシグナル伝達が通るように、広範なGタンパク質(Offermanns, S. および M. I. Simon (1995) J. Biol. Chem. 270 : 15175 - 15180)との結合が実証されたGタンパク質である $G_{15/16}$ とGCRECを同時発現させることができる。或いは、GCRECを内在性GPCRが存在しない遺伝子組換え酵母系に発現させて、GCREC活性化スクリーニングにおける有利なバックグラウンドを提供することができる。これらの酵母系において、ヒトGPCRおよびGタンパク質を内在性酵母フェロモン受容体経路の対応する成分と置換する。また、下流のシグナル伝達経路を変更して、シグナルに対する正常な酵母の応答が、選択培地における正の成長或いはレポーター遺伝子の発現に変換されるようにする(Broach, J. R. および J. Thorner (1996) Nature 384 (supp.) : 14 - 16)。受容体を、既知のGPCRリガンド

30

40

50

およびその他の天然の生理活性分子を含む推定リガンドに対してスクリーニングする。

【0294】

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うことができるであろう。特定の好適な実施例に基づいて本発明を説明したが、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、分子生物学或いは関連する分野の専門家には明らかな、本明細書に記載の本発明の実施例の様々な改変は、特許請求の範囲に含まれる。

【0295】

(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に対する系統的な名称を示す。 10

【0296】

表2は、本発明のポリペプチドに最も近いGenBankの相同体のGenBankの識別番号およびアノテーションを示す。ポリペプチドとそのGenBankの相同体との間の一致を表す確率値スコアも示す。

【0297】

表3は、推定上のモチーフおよびドメインを含むポリペプチド配列の構造的な特徴、並びにポリペプチドの分析に用いた方法、アルゴリズム、および検索可能なデータベースを示す。

【0298】

表4は、ポリヌクレオチド配列の組み立てに用いたcDNA断片およびゲノムDNA断片のリスト、並びにポリヌクレオチド配列の選択された断片のリストを示す。 20

【0299】

表5は、本発明のポリヌクレオチドの代表的なcDNAライブラリを示す。

【0300】

表6は、表5に示すcDNAライブラリの作製に用いた組織およびベクターを示す付録である。

【0301】

表7は、本発明のポリヌクレオチド及びポリペプチドの分析に用いたツール、プログラム、及びアルゴリズム、並びにその説明、引用文献、閾値パラメータを示す。 30

【表1】

表 1

インサイト プロジェクトID	ポリベブチド SEQ ID NO:	インサイト ポリペプチドID	ポリヌクレオチド SEQ ID NO:	インサイト ポリヌクレオチドID
536482	1	536482CD1	22	536482CB1
1316020	2	1316020CD1	23	1316020CB1
2816437	3	2816437CD1	24	2816437CB1
2289894	4	2289894CD1	25	2289894CB1
7066050	5	7066050CD1	26	7066050CB1
5376785	6	5376785CD1	27	5376785CB1
3082743	7	3082743CD1	28	3082743CB1
7472361	8	7472361CD1	29	7472361CB1
7472363	9	7472363CD1	30	7472363CB1
7472364	10	7472364CD1	31	7472364CB1
7472434	11	7472434CD1	32	7472434CB1
7472435	12	7472435CD1	33	7472435CB1
7472438	13	7472438CD1	34	7472438CB1
7472439	14	7472439CD1	35	7472439CB1
7472440	15	7472440CD1	36	7472440CB1
7472443	16	7472443CD1	37	7472443CB1
7472445	17	7472445CD1	38	7472445CB1
7472446	18	7472446CD1	39	7472446CB1
7472451	19	7472451CD1	40	7472451CB1
7472456	20	7472456CD1	41	7472456CB1
7472457	21	7472457CD1	42	7472457CB1

10

20

30

40

【表 2】

表2-1

ポリペプチド SEQ ID NO:	インサイト ポリペプチドID	GenBank ID NO:	確率スコア	GenBank 相同体
2	1316020	g927209	4.10E-21	α 1Cアドレナリン作用レセプターインフォーム2 [ヒト] (Hirasawa, A., 他 (1995) Cloning, functional expression and tissue distribution of human alpha 1c-adrenoceptor splice variants. FEBS Lett. 363 (3), 256-260.)
4	2289894	g992582	3.20E-35	Gタンパク質結合7回膜貫通受容体 [メダカ] (Yasuoka, A., 他 (1995) Molecular cloning of a fish gene encoding a novel seven-transmembrane receptor related distantly to catecholamine, histamine, and serotonin receptors. Biochim. Biophys. Acta 1235, 467-469.)
5	7066050	g6531395	2.50E-175	成長因子調節Gタンパク質結合受容体Nrg-1 [ドブネズミ] (Glickman, M., 他 (1999) Molecular cloning, tissue-specific expression, and chromosomal localization of a novel nerve growth factor- regulated G-protein-coupled receptor, nrg-1. Mol. Cell. Neurosci. 14, 141-152.)
6	5376785	g460318	5.30E-72	Gタンパク質結合受容体 [ハツカネズミ] (Harrigan, M.T., 他 (1991) Identification of a gene induced by glucocorticoids in murine T-cells: a potential G protein-coupled receptor. Mol. Endocrinol. 5: 1331-1338.)

10

20

30

40

表2-2

ポリペプチド SEQ ID NO:	インサイト ポリペプチドID	GenBank ID NO:	確率スコア	GenBank 相同体
7	3082743	g6178006	5.50E-141	臭気物質受容体 MOR83 [ハツカネズミ] (Tsuboi,A.他 (1999) Olfactory neurons expressing closely linked and homologous odorant receptor genes tend to project their axons to neighboring glomeruli on the olfactory bulb. J. Neurosci. 19:8409-8418.)
8	7472361	g3927808	1.80E-85	嗅覚受容体様タンパク質COR3'β [ニワトリ] (Reitman,M.他 (1993) Primary sequence, evolution, and repetitive elements of the Gallus gallus (chicken) beta-globin cluster. Genomics 18: 616-626.)
9	7472363	g6532001	1.90E-90	臭気物質受容体 S19 [ハツカネズミ]
10	7472364	g4680268	2.00E-95	臭気物質受容体 S46 [ハツカネズミ] (Mainic,B.他 (1999) Combinatorial receptor codes for odors. Cell 96: 713-723.)
11	7472434	g5869927	2.20E-98	嗅覚受容体 [ハツカネズミ]
12	7472435	g5869916	4.00E-138	嗅覚受容体 [ハツカネズミ]
13	7472438	g9963968	1.00E-141	臭気物質受容体 M72 [ハツカネズミ] (Zheng,C.他 (2000) Peripheral olfactory projections are differentially affected in mice deficient in a cyclic nucleotide-gated channel subunit. Neuron 26: 81-91.)

【表4】

表2-3

ポリペプチド SEQ ID NO:	インサイト ポリペプチドID	GenBank ID NO:	確率スコア	GenBank 相合体
14	7472439	g11692541	1.00E-123	臭気物質受容体 K23 [ハツカネズミ] (Xie,S.Y.他 (2000) Characterization of a cluster comprising approximately 100 odorant receptor genes in mous. Mamm. Genome 11: 1070-1078.)
15	7472440	g11692535	1.00E-135	臭気物質受容体 K21 [ハツカネズミ] (Xie,S.Y.他 (2000) Characterization of a cluster comprising approximately 100 odorant receptor genes in mous. Mamm. Genome 11: 1070-1078.)
16	7472443	g1246534	3.50E-82	嗅覚受容体 4 [ニワトリ] (Leibovici,M.他 (1996) Avian olfactory receptors: differentiation of olfactory neurons under normal and experimental conditions. Dev. Biol. 175:118-131.)
17	7472445	g5453092	1.00E-102	嗅覚受容体 [ハツカネズミ] (Rouquier,S.他 (2000) The olfactory receptor gene repertoire in primates and mouse: evidence for reduction of the functional fraction in primates. Proc. Natl. Acad. Sci. U.S.A. 97: 2870-2874.)
18	7472446	g3983382	4.00E-65	嗅覚受容体 E3 [ハツカネズミ] (Krautwurst,D.他 (1998) Identification of ligands for olfactory receptors by functional expression of a receptor library. Cell 95: 917-926)

10

20

30

40

表2-4

ポリペプチド SEQ ID NO:	インサイト ポリペプチドID	GenBank ID NO:	確率スコア	GenBank 相同体
19	7472451	g1256393	1.90E-97	味覚受容体タンパク質 TB 641 [ドブネズミ] (Thomas, M.B. 他 (1996) Chemoreceptors expressed in taste, olfactory and male reproductive tissues. Gene 178: 1-5)
20	7472456	g33983374	2.80E-82	嗅覚受容体 C6 [ハツカネズミ] (Krautwurst, D. 他 (1998) Identification of ligands for olfactory receptors by functional expression of a receptor library. Cell 95: 917-926.)
21	7472457	g6178006	3.90E-92	臭気物質受容体 MOR83 [ハツカネズミ] (Tsuboi, A. 他 (1999) Olfactory neurons expressing closely linked and homologous odorant receptor genes tend to project their axons to neighboring glomeruli on the olfactory bulb. J. Neurosci. 19:8409-8418.)

10

20

30

40

表3-1

SEQ ID NO:	インサイトポリペプチド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(signature)配列, ドメイン及びモチーフ	分析方法及びデータベース
1	536482CD1	99	T34		膜貫通ドメイン: G50-L70 バクテリアアロドプシンシグネチャ: Y32-K83 シグナルペプチド: M1-S57 G-タンパク質結合受容体ファミリー2 シグネチャ: E54-G119 TA2R_HUMAN, β イソホーミング PD168643: S70-L135 (P値 = 4.3e-07)	HMMER PROFILESCAN SPSCAN PROFILESCAN BLAST-PRODOM SPSCAN
2	1316020CD1	139	S14 T61 S131		シグナルペプチド: M1-S48 メラノコルチン受容体ファミリーシグネチャ: L39-F56 バクテリシリンV2受容体シグネチャ: F56-S75	SPSCAN BLIMPS-PRINTS BLIMPS-PRINTS
3	2816437CD1	82	T4 S58 S21 S75 S78	N46	膜貫通ドメイン: L18-P43 7回膜貫通受容体(ロドプシンファミリー) ドメイン: G31-Y290 G-タンパク質結合受容体シグネチャ BL00237: V81-P120; F181-Y192; L234-A260 トロポキサン受容体シグネチャ PR00429: A199-L220	HMMER HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS
4	2289994CD1	368	S305	N3 N8	G-タンパク質結合受容体 DM00013(P25962)31-357: A13-R201; P206-Y290	BLAST-DOMO

【表7】

10

20

30

40

表3-2

SEQ ID NO.	インサイトポリペプチド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(signature)配列, ドメイン及びモチーフ	分析方法及びデータベース
5	7066050CD1	398	T79 T309 S340 S361 T22 T100 S146 S237 S363	N20	膜貫通ドメイン: V42-V60; V194-Y212 7回膜貫通受容体(ロドプシンファミリー) ドメイン: E53-Y306 G-タンパク質結合受容体シグネチャ BL00237: L101-R140; P244-L270; N298-R314 ロドプシン様GPCRスーパーファミリーシグネチャ PR00237: A38-G62; M71-N92; V115-M137 R150-L171; Y193-Y216; L249-L273 A288-R314 EDG1オーファン受容体シグネチャPR00642: E13-R29; V60-L74; S96-V115 D274-F291 G-タンパク質結合受容体 DM00013(P21453)39-326: L36-V321 G-タンパク質結合受容体モチーフ: A121-M137	HMMER HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-DOMO MOTIFS

10

20

30

40

表3-3

SEQ ID NO:	インサイトポリペプチド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	シグネチヤ(signature)配列, ドメイン及びモチーフ	分析方法及びデータベース
6	5376785CD1	153	T20, S46, S110, S116, S144		G-タンパク質結合受容体 DM00013(P30731)65-36142-S91 T細胞プレカーサー-グルコルチコイド誘導Gタンパク質結合膜貫通糖タンパク質シグナル補体スプライシングからの推定G-タンパク質結合受容体: PD061453:C76-S153	BLAST-DMC BLAST-PRODOM
					シグナル分割部位: M1-C40	SPSCAN
					膜貫通ドメイン: T20-L38	EMMER
					7回膜貫通受容体-(ロドプシンファミリー): A12-Y75	EMMER-PFAM
					視覚色素(オプシン)網膜結合部位: W62-R103	PROFILESAN
					G-タンパク質結合受容体: BL00237C:R15-Y41; BL00237D:S67-R83	BLIMPS-BLOCKS
					ロドプシン様GPCRスーパーファミリー PR00237A:M23-S47; PR00237E:L22-L45; PR00237F:T20-L44; PR00237G:Y57-R83	BLIMPS-PRINTS
					推定Gタンパク質結合受容体ドメイン: PR010181:C76-L89	BLIMPS-PRINTS

10

20

30

40

表3-4

SEQ ID NO:	インサイトポリペプチド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(signature)配列、ドメイン及びモチーフ	分析方法及びデータベース
7	3082743CD1	313	S229 S67 T259 T163 S224 T288	N5 N65	膜貫通ドメイン: M29-A48; G100-V118; Y193-A213 7回膜貫通受容体(ロドプシンファミリー) ドメイン: G41-Y287 G-タンパク質結合受容体シグネチャ BL00237: K90-P129; T279-K295 G-タンパク質結合受容体シグネチャ: F102-L148 ロドプシン様GPCRスーパーファミリーシグネチャ PR00237: L26-V50; M59-K80; L104-I126 A236-R260; K289-L295 嗅覚受容体シグネチャPR00245: M59-K80; Y177-D191; L237-G252 V271-L282; T288-R302 嗅覚受容体タンパク質受容体様 Gタンパク質結合膜貫通糖タンパク質 多重塩基子ファミリー PD149621: V246-Q305 Gタンパク質結合受容体 DM00013(S29710)15-301: F28-L301 Gタンパク質結合受容体モチーフ: A110-I126	HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILESAN BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO MOTIFS

10

20

30

40

表3-5

SEQ ID NO:	インサイト ポリバプチド 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(signature)配列, ドメイン及びモチーフ	分析方法及び データベース
8	7472361CD1	315	S14 T31 S32 T101 S114 T155 S224 T63 T307 T354 Y351	N12 N61 N89	膜貫通ドメイン: F78-I96; I182-I210; I239-I262 7回膜貫通受容体(ロドプシンファミリー) ドメイン: G88-I197; F252-Y338 Gタンパク質結合受容体シグネチャ BL00237: H137-P176; F253-Y264; D278-S304 P330-R346 嗅覚受容体シグネチャPR00245: M106-R127; S224-N238; L284-I299 推定Gタンパク質結合受容体RA1C PD170483: V291-F353 Gタンパク質結合受容体 DM00013(G45774)18-309: V81-E347 Gタンパク質結合受容体モチーフ: M157-I173	HMMER HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO MOTIFS

10

20

30

40

表3-6

SEQ ID NO:	インサイトポリペプチド 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(signature)配列、ドメイン及びモチーフ	分析方法及びデータベース
9	7472363CD1	356	T110 S190 T226 S232 S295	N5 N44	膜貫通ドメイン: Y37-S55; P60-L84; V204-I223 7回膜貫通受容体(ロドプシンファミリー) ドメイン: G43-Y294 Gタンパク質結合受容体シグネチャ BL00237: P92-P131; P286-R302 Gタンパク質結合受容体シグネチャ: F104-S154 嗅覚受容体シグネチャPR00245: M61-T82; A179-D193; L240-V255 推定Gタンパク質結合受容体RA1C PD170483: I247-F309 Gタンパク質結合受容体 DM00013(S29707)18-306: E23-R302 Gタンパク質結合受容体モチーフ: L112-I128	HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILESAN BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO MOTIFS

10

20

30

40

表3-7

SEQ ID NO:	インサイトポリペプチド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V(signature)配列, ドメイン及びモチーフ	分析方法及びデータベース
10	7472364CD1	311	T56 S69 T110 N5 N44 S179 T262 T309		膜貫通ドメイン: F33-I51; I137-I165; I194-I217 7回膜貫通受容体(ロドプシンファミリー) ドメイン: G43-I152; F207-Y293 Gタンパク質結合受容体シグネチヤ BL00237: H92-P131; F208-Y219; D233-S259 P285-R301 嗅覚受容体シグネチヤPR00245: M61-R82; S179-N193; L239-I254 推定Gタンパク質結合受容体RA1C PD170483: V246-F308 Gタンパク質結合受容体 DM00013(G45774)18-309: P20-E302 Gタンパク質結合受容体モチーフ: M112-I128	HMMER HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO MOTIFS

10

20

30

40

表3-8

SEQ ID NO:	インサイトポリペプチド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(signature)配列, ドメイン及びモチーフ	分析方法及びデータベース
11	7472434CD1	354	S340 S28 S86 N93 S95 S3 T60 S64 T116 S316 S336	潜在的グリコシル化部位	膜貫通ドメイン: I218-A235 7回膜貫通受容体(ロドプシンファミリー) ドメイン: N93-Y315 Gタンパク質結合受容体シグネチャ BL00237: R118-P157; T233-Y244; T307-K323 Gタンパク質結合受容体シグネチャ: Y130-L175 ロドプシン様GPCRスーパーファミリーシグネチャ PR00237: T132-I154; L225-L248; K297-K323 嗅覚受容体シグネチャPR00245: N87-L108; Y203-D217; F264-G279 V299-L310; S316-L330 受容体嗅覚タンパク質受容体様 Gタンパク質結合膜貫通糖タンパク質 多重伝子ファミリー-PD000921: L194-H270 Gタンパク質結合受容体 DM00013(P23275)17-306: N92-L330 Gタンパク質結合受容体モチーフ: T138-I154	HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILES SCAN BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO MOTIFS

10

20

30

40

【表 1 4】

SEQ ID NO:	インサイトポリペプチド 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(signature)配列, ドメイン及びモチーフ	分析方法及びデータベース
12	7472435CD1	319	T50 S68 T88 S277 S298	N5 N66	シグナルペプチド: M1-G42 膜貫通ドメイン: Y30-L48; M60-L83; I198-T225 7回膜貫通受容体(ロドプシンファミリー) ドメイン: G42-Y297 Gタンパク質結合受容体シグネチャ BL00237: E91-P130; T289-K305 Gタンパク質結合受容体シグネチャ: L104-A147 嗅覚受容体シグネチャPR00245: M60-L81; F178-D192; F239-G254 I281-L292; S298-L312 受容体嗅覚タンパク質受容体様 Gタンパク質結合膜貫通糖タンパク質 多重伝子ファミリー-PD000921: L167-L246 Gタンパク質結合受容体 DM00013(P23276)17-306: L18-L312 Gタンパク質結合受容体モチーフ: T111-I127	SPSCAN HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILESAN BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO MOTIFS

10

20

30

40

【表 1 5】

表3-10

SEQ ID NO:	インサイトポリペプチド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(signature)配列, ドメイン及びモチーフ	分析方法及びデータベース
13	7472438CD1	309	T8 S67 S156 S190 S228 T18 T78 S87 S290 T303	N5 N65	膜貫通ドメイン: F28-L48; M98-D121 7回膜貫通受容体(ロドプシンファミリー) ドメイン: G41-Y289 Gタンパク質結合受容体シグネチャ BL00237: N90-P129; I281-K297 Gタンパク質結合受容体シグネチャ: Y102-A150 嗅覚受容体シグネチャPR00245: M59-K80; Y176-S190; F237-G252 A273-L284; S290-L304 受容体嗅覚タンパク質受容体様 Gタンパク質結合膜貫通糖タンパク質 多重遺伝子ファミリー-PD000921: L166-L244 Gタンパク質結合受容体 DM00013(S51356)18-307: L17-V300	HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILESAN BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO

10

20

30

40

表3-11

SEQ ID NO:	インサイト 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(signature)配列, トメイン及びモチーフ	分析方法及び データベース
14	7472439CD1	310	S7 S66 S228 T77 S86 S290	N5 N64 N164 N189 N227	膜貫通ドメイン: L26-I45; M58-T77; P128-L145 E195-I220 7回膜貫通受容体(ロドプシンファミリー) ドメイン: G40-Y289 6タンパク質結合受容体シグネチャ BL00237: N89-P128; V281-K297 6タンパク質結合受容体シグネチャ: F101-G151 ロドプシン様GPCRスーパーファミリーシグネチャ PR00237: P25-G49; M58-K79; F103-I125 V198-L221; K271-K297 嗅覚受容体シグネチャ PR00245: M58-K79; F176-S190; F237-G252 S273-L284; S290-L304 受容体嗅覚タンパク質受容体様 6タンパク質結合膜貫通種タンパク質 多重塩基子ファミリー PD000921: L165-I245 6タンパク質結合受容体 DM00013(S29709)11-299: T17-G305 6タンパク質結合受容体モチーフ: S109-I125	HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILES SCAN BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO MOTIFS

10

20

30

40

【表 17】

表3-12

SEQ ID NO:	インサイトポリペプチド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(signature)配列、ドメイン及びモチーフ	分析方法及びデータベース
15	7472440CD1	311	S91 S67 T78 S291	N5 N65	膜貫通ドメイン: F29-T47; Q100-M118; P129-L146 7回膜貫通受容体(ロドプシンファミリー) ドメイン: G41-Y290 Gタンパク質結合受容体シグネチャ BL00237: N90-P129; V282-K298 Gタンパク質結合受容体シグネチャ: F102-G150 ロドプシン様GPCRスーパーファミリーシグネチャ PR00237: P26-R50; M59-K80; F104-I126 V199-I222; K272-K298 嗅覚受容体シグネチャ PR00245: M59-K80; Y177-S191; F238-G253 S274-L285; S291-L305 受容体嗅覚タンパク質受容体様 Gタンパク質結合膜貫通タンパク質 多重伝達子ファミリー PD000921: V166-I246 Gタンパク質結合受容体 DM00013(S29709)I1-299: T18-L305 Gタンパク質結合受容体モチーフ: S110-I126	HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILES CAN BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODOM BLAST-DMO MOTIFS

10

20

30

40

表3-13

SEQ ID NO:	インサイトポリペプチド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(signature)配列, ドメイン及びモチーフ	分析方法及びデータベース
16	7472443CD1	314	S67 S267 T87 S264 S291	N5 N210	膜貫通ドメイン: L29-Y56; F101-D121; I197-I221 H244-S262 7回膜貫通受容体(ロドプシンファミリー) ドメイン: G41-Y290 Gタンパク質結合受容体シグネチャ BL00237: T90-P129; I282-K298 Gタンパク質結合受容体シグネチャ: F102-V150 嗅覚受容体シグネチャPR00245: M59-Q80; F177-D191; F238-G253 V274-I285; S291-A305 受容体嗅覚タンパク質受容体様 Gタンパク質結合膜貫通タンパク質 多重伝子ファミリー PD000921: L166-I245 Gタンパク質結合受容体 DN00013(S51356)18-307: P21-A300 Gタンパク質結合受容体モチーフ: T110-I126 ATP/GTP 結合部位(ループ) A83-T90	HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILESAN BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMC MOTIFS MOTIFS

10

20

30

40

表3-14

SEQ ID NO:	インサイトポリペプチド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(signature)配列、ドメイン及びモチーフ	分析方法及びデータベース
17	7472445CD1	346	T123 S32 S99 N37 N97 S188 T298 T25 T169 S177 S323 S343	N37 N97	膜貫通ドメイン: I61-L83; I133-D153; M229-L247 7回膜貫通受容体(ロドプシンファミリー) ドメイン: G73-Y322 Gタンパク質結合受容体シグネチャ BL00237: K122-F161; T314-K330 Gタンパク質結合受容体シグネチャ: I134-F178 ロドプシン様GPCRスーパーファミリーシグネチャ PR00237: L58-F82; M91-Q112; F136-V158 I231-I254; K304-K330 嗅覚受容体シグネチャ PR00245: M91-Q112; Y209-D223; F270-G285; I306-L317; S323-V337 受容体嗅覚タンパク質受容体様 Gタンパク質結合膜貫通タンパク質 多重伝子ファミリー PD149621: V279-K340 Gタンパク質結合受容体 DM00013(P23276)17-306: Q56-G338 Gタンパク質結合受容体モチーフ: T142-V158	HMMER HMMER-PPAM BLIMPS-BLOCKS PROFILESAN BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO MOTIFS

10

20

30

40

表3-15

SEQ ID NO:	インサイトポリペプチド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	シグネチヤ(signature)配列, ドメイン及びモチーフ	分析方法及びデータベース
18	7472446CD1	316	S67 S193 T268 T78 S137 T192 S267 S291	N5 N19	膜貫通ドメイン: L34-A53; E196-Y218 7回膜貫通受容体(ロドプシンファミリー) ドメイン: S41-Y290 Gタンパク質結合受容体シグネチヤ BL00237: N90-P129; T282-M298 Gタンパク質結合受容体シグネチヤ: F102-T147 ロドプシン様GPCRスーパーファミリーシグネチヤ PR00237: L26-T50; M59-K80; A104-I126 A140-V161; V199-L222; A237-L261 N272-M298 嗅覚受容体シグネチヤ PR00245: M59-K80; L177-D191; L238-G253 I274-L285; S291-L305 受容体嗅覚タンパク質受容体様 Gタンパク質結合膜貫通糖タンパク質 多重伝子ファミリー PD149621: T246-K307 Gタンパク質結合受容体 DM00013 (P23275) 17-306: A29-G306	HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILESCAN BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODOM BLAST-DMO

10

20

30

40

表3-16

SEQ ID NO:	インサイト 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(signature)配列, ドメイン及びモチーフ	分析方法及び データベース
19	7472451CD1	453	S81 T213 T331 T151	N19 N441	シグナルペプチド: M1-G55 膜貫通ドメイン: L43-I60; I217-V236 7回膜貫通受容体(ロドプシンファミリー) ドメイン: G55-I263 Gタンパク質結合受容体シグネチャ BL00237: R104-P143; T294-K310 Gタンパク質結合受容体シグネチャ: F116-T162 嗅覚受容体シグネチャ PR00245: M73-K94; I191-D205; F252-V267 V286-L297; T303-G317 受容体嗅覚タンパク質受容体様 Gタンパク質結合膜貫通シグナル 多重膜伝子ファミリー PD000921: L180-L259 Gタンパク質結合受容体 DN00013 (S29710) I5-301: L41-L316	SPSCAN HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILESCAN BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO

10

20

30

40

【表 2 2】

表3-17

SEQ ID NO:	インサイト 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(signature)配列, ドメイン及びモチーフ	分析方法及び データベース
20	7472456CD1	323	S49 S67 S188 T230 T40 T291 S320	N5 N65	膜貫通ドメイン: F25-I45; I57-I85; Y102-D121 L194-Y218 7回膜貫通ドメイン受容体(ロドプシンファミリー) ドメイン: G41-F290 Gタンパク質結合受容体シグネチャ BL00237: H90-P129; T282-Q298 Gタンパク質結合受容体シグネチャ: Y102-A147 ロドプシン様GPCRスーパーファミリー-シグネチャ PR00237: M59-K80; Y104-I126; A199-L222 K272-Q298 嗅覚受容体シグネチャ PR00245: M59-K80; F177-D191; F238-G253 A274-L285; T291-L305 受容体嗅覚タンパク質受容体様 Gタンパク質結合膜貫通糖タンパク質 多重伝子ファミリー-PD000921: L166-L245 Gタンパク質結合受容体 DM00013(P23267)20-309: F17-L305 Gタンパク質結合受容体モチーフ: T110-I126	HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILESCAN BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODOM BLAST-DMO MOTIFS

10

20

30

40

表3-18

SEQ ID NO:	インサイトポリペプチド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	シグネチヤ(signature)配列, ドメイン及びモチーフ	分析方法及びデータベース
21	7472457CD1	318	S67 T288	N5 N65 N137	膜貫通ドメイン: V30-V49 7回膜貫通受容体(ロドプシンファミリー) ドメイン: G41-Y287 Gタンパク質結合受容体シグネチヤ BL00237: P90-P129; T279-I295 Gタンパク質結合受容体シグネチヤ: F102-A147 ロドプシン様GPCRスーパーファミリーシグネチヤ PR00237: V26-T50; M59-R80; F104-I126 V140-A161; M199-L222; A236-R260 K269-I295 嗅覚受容体シグネチヤ PR00245: M59-R80; F177-D191; L237-V252 V271-L282; T288-W302 受容体嗅覚タンパク質受容体様 Gタンパク質結合膜貫通タンパク質 多重置換子ファミリー PD000921: L166-I244 Gタンパク質結合受容体 DM00013(S29710)15-301: L17-W302	HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILESAN BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO

10

20

30

40

表4-1

ポリヌクレオチド SEQ. ID NO:	インサイト ポリヌクレオチドID	配列長さ	選択された断片	配列断片	5'位置	3'位置
22	536482CB1	1348	1-174, 812-868	6871485H1 (BRAGNON02) 1347882T6 (PROSN0T11) 536482R6 (LNODN0T02) 708761LH1 (BRAUTDR03)	218 457 1 798	950 1058 436 1348
23	1316020CB1	1446	1-750, 851-930, 771-832	5599305F8 (UTREN0N03) 1836450R6 (BRAINO01) 6442433H1 (BRAENO02) 7121815H1 (BRAHNOE01) 5174595H1 (EPIBXT01) 5575376H1 (BRAPNO04)	716 1174 833 1 325 508	1330 1446 1440 445 619 769
24	2816437CB1	1463	155-231, 1427-1463, 644-800	2780847F6 (OVARUT03) 2816437H1 (BRSTNO14) 70171099V1 6915195M1 (PITUDIR01) 70173319V1 2943542H2 (BRAITUT23) FL2289894_00001	250 1 664 880 723 585 63	716 276 1054 1463 1225 722 1435
25	2289894CB1	1435	477-681, 1289-1435	95743982	1	455
26	7066050CB1	2147	629-1512, 1-141	7066050H1 (BRAINOR01) FL7066050_00001 70700621V1	1 11 1617	512 2132 2147
27	5376785CB1	1989	1493-1989, 1-53, 875-990, 157-485, 1338-1432, 1095-1240	7101536H1 (BRAWTDR02) 71217673V1 70822278V1 70818975V1 71217707V1	712 1242 1 985 485	1256 1989 602 1467 1122
28	3082743CB1	942	897-942	GBI.g2121229.raw FL3082743_00001	1 121	360 942
29	7472361CB1	948	577-948	GNN.g6671985_006	1	948
30	7472363CB1	1071	554-1071	GNN.g6671985_018	1	1071

【表 2 5】

10

20

30

40

表4-2

ポリスクレオチド SEQ ID NO:	インサイト ポリスクレオチドID	配列長さ	選択された断片	配列断片	5'位置	3'位置
31	7472364CB1	1001	1-55, 435-1001	GNN.g6671985_022	1	1001
32	7472434CB1	1065	1-356, 1002-1065, 561-594	GNN.g6630753_006	1	1065
33	7472435CB1	963	1-27, 762-963	GNN.g6630753_008	1	963
34	7472438CB1	1101	1-138, 582-712	GNN.g6635275_002	1	1101
35	7472439CB1	933		GNN.g6635275_006	1	933
36	7472440CB1	936	425-544	GNN.g6635275_018	1	936
37	7472443CB1	945	922-945	GNN.g6642708_020	1	945
38	7472445CB1	1041	1008-1041, 81-107	GNN.g6648400_008	1	1041
39	7472446CB1	951	920-951, 578-607	GNN.g6648400_012	1	951
40	7472451CB1	1395	1-87, 1208-1395, 857-1059	GNN.g6648431_006	1	1395
41	7472456CB1	972	462-972	GNN.g6648431_016	1	972
42	7472457CB1	957	916-957	GNN.g6648431_018	1	957

10

20

30

40

【表 2 6】

表5

ポリヌクレオチド SEQ ID NO:	インサイト プロジェクトID	代表的ライブラリ
22	536482CB1	PROSNOT11
23	1316020CB1	BRAINNOY02
24	2816437CB1	OVARFUT03
25	2289894CB1	BRAINNO01
26	7066050CB1	LEUKNOT02
27	5376785CB1	BRAXNOT01

【表 2 7】

10

20

30

ライブラリ	ベクター	ライブラリの説明
OVRTUT03	pINCY	ライブラリは、52歳の混血女性の腔式子宮全摘出術、両卵管卵巣摘出術、腹膜及びびリンパ組織の生検、局所的なリンパ節切除、及び腹膜組織の破壊の際に、左卵巣から卵巣腫瘍組織を採取し、その卵巣腫瘍組織から単離したRNAを用いて作製した。病理学的には、左卵巣に腫瘤を形成しているグレード3(4)の漿液性未分化(seroanaplastic)浸潤癌である。左卵巣と卵管、右卵巣と卵管、子宮の後面、及び盲管の表面上に複数の腫瘍が散在していた。子宮内膜は萎縮していた。複数(2)の平滑筋腫が確認され、1つは漿膜下で、1つは壁内であった。さらに、病変は大網、盲管腹膜、左広間膜腹膜及び結腸間膜を含むような転移グレード3の漿液性未分化癌を示した。患者の病歴には、乳癌及び慢性消化性潰瘍、関節痛がある。家族歴には、大腸癌及び脳血管障害、乳癌、II型糖尿病、食道癌、抑うつ障害が含まれる。
PROSNOT11	pINCY	ライブラリは、自傷銃創で死亡した28歳白人男性の前立腺組織から単離したRNAを用いて作製した。

【表 2 8】

10

20

30

40

表6-2

ライブラリ	ベクター	ライブラリの説明
BRA1IN001	PSPORT1	ライブラリは、26歳の白人男性の頭蓋形成術及び大脳髄膜病変の切除の際に採取した脳組織から作られたRNAからの独立したクローン4,88×10 ⁶ 個から作製され標準化された。関連した腫瘍組織は病理学的には、右前頭頂のグレード4の乏星状細胞腫 (oligoastrocytoma) である。
BRA1IN002	pINCY	この大きくサイズ分割され、標準化されたライブラリは、心不全で死亡した35歳の白人男性から採取した中脳、下外側皮質、髄質、及び部分的な後部皮質の組織から単離した mRNA より作られたプールされた cDNA を用いて作製された。病理学的には、大脳新皮質の中等度の軟膜線維症及び多発性微小梗塞を示していた。微視的には、大脳半球は限局性石灰化を伴なう中程度の軟膜の線維症に罹っていた。収縮の証拠及び僅かに好酸性の錐体ニューロンが大脳半球全域にあった。更に、大脳半球全域に散乱して、神経膠症に取り囲まれたキャピタリオンが多発性極微小領域が存在していた。患者の病歴には、拡張型心筋症、うっ血性心不全、心臓肥大、脾腫及び肝臓肥大が含まれる。このサイズ選択されたライブラリからの 2.8×10 ⁵ の独立したクローンは著しく長い(48時間/回) アニリングハイブリダイゼーションが用いられたの以外は Soares 他、PNAS(1994) 91:9928-9232 及び Bonaldo 他、Genome Research 6 (1996):791 の条件に従って2回標準化した。
BRAXN001	pINCY	ライブラリは、70歳の男性の小脳組織から単離した RNA を用いて作製した。患者の病歴には、慢性閉塞性航空障害及び左心不全が含まれる。
LEUKN002	pINCY	ライブラリは、血液型O+の45歳の女性の白血球より単離した RNA を用いて作製された。ドナーはサイトメガロウイルス (CMV) に対して陽性であった。

【表 29】

表7-1

プログラム名	説明	引用文献	パラメーター関連値
ABI FACTURA	核酸配列においてベクター配列を除去して不定の塩基をマスクするプログラム。	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	Fast Data Finder は、アミノ酸または核酸配列の比較及び注釈付け (annotation) に有用である。	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	不一致<50%
ABI AutoAssembler	核酸配列を構築するプログラム。	Applied Biosystems, Foster City, CA.	
BLAST	Basic Local Alignment Search Tool は、アミノ酸及び核酸配列の配列類似性検索に有用であり、blastp 及び blastn, blastx, tblastn, tblastx の5つのファンクションがある。	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	Pearson 及び Lipman アルゴリズムは、問合わせの配列と同種の配列群との類似性を検索する。FASTA は、fasta 及び tfasta, fastx, tfastx, ssearch の少なくとも5つのファンクションを含む。	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; 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	Blocks IMPROVED Searcher は、BLOCKS 及び PRINTS、DOMO、PRODOM、PFAM データベースにおける配列に対して、ある配列の一致性を調べ、遺伝子ファミリー及び配列相同性、構造的フィンガープリント領域を探索する。	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; Artwood, T.K. 他 (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	確率値=1.0E-3 以下
HMMER	PFAM などのタンパク質ファミリーコンセンサス配列の隠れマルコフモデル(HMM)に基づいたデータベースに対して問合せ配列を検索するアルゴリズム。	Krogh, A. 他 (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. 他 (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

プログラム名	説明	引用文献	パラメーター閾値
ProfileScan	Prosites で定義された配列パターンと一致するタンパク質配列における構造及び配列のモチーフを検索するアルゴリズム。	Gribskov, M. 他 (1988) CABIOS 4:61-66; Gribskov, 他 (1989) Methods Enzymol. 183:146-159; Bairoch, A. 他 (1997) Nucleic Acids Res. 25:217-221	標準化された質のスコア≧特定の Prositesモチーフに対する GCG 指定 "HIGH" 値 通常、スコア=1.4-2.1
Phred	高い感度及び確率で自動配列決定機のトレースを調べる塩基読み出しアルゴリズムである。	Ewing, B. 他 (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	Smith-Waterman アルゴリズムの効率的なインプリメンテーションに基づく SWAT や CrossMatch を含む Phils Revised Assembly プログラムであって、配列相同性の検索及び DNA 配列の構築に有用である。	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; Green, P., University of Washington, Seattle, WA.	スコア=120 以上 一致長さ=56 以上
Consed	Phrap で構築したものの表示及び編集をするためのグラフィックツールである。	Gordon, D. 他 (1998) Genome Res. 8:195-202.	
SPScan	タンパク質配列をスキッピングして分泌シグナルペプチドの存在を調べる重み付けマトリクス解析プログラムである。	Nielson, H. 他 (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	スコア=3.5 以上
TMAP	蛋白質配列での膜貫通セグメントの描写及び配向の決定のための、加重マトリクスを用いたプログラムである。	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	蛋白質配列での膜貫通セグメントの描写及び配向の決定のための、隠された Markov モデル(HMM)を用いたプログラムである。	Sonnhammer, E.L. 他 (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et 他., es., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	Prosites で定義された配列と一致したパターンについてアミノ酸配列を検索するプログラムである。	Bairoch 他, (1997) Nucleic Acid Res. 25: 217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

【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
13 September 2001 (13.09.2001)

PCT

(10) International Publication Number
WO 01/66742 A2

(51) International Patent Classification: C12N 15/12, 15/10, 15/66, C07K 14/705, 16/28, A01K 67/027, A61K 38/17, C12Q 1/68, G01N 33/68

(21) International Application Number: PCT/US01/06814

(22) International Filing Date: 1 March 2001 (01.03.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/186,854 3 March 2000 (03.03.2000) US
60/188,384 10 March 2000 (10.03.2000) US
60/190,153 17 March 2000 (17.03.2000) US
60/190,730 29 March 2000 (29.03.2000) US

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

(72) Inventors; and
(75) Inventors/Applicants (for US only): I.A.I., Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US); TANG, Y. Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US); PATTERSON, Chandra [US/US]; 490 Sherwood Way Rt. Menlo Park, CA 94025 (US); YAO, Monique, G. [US/US]; 111 Frederick Court, Mountain View, CA 94025 (US); SIIIH, Leo, L. [US/US]; 1081 Lantana Drive, Apt. B, Palo Alto, CA 94305 (US); TRIBOULEY, Catherine, M. [FR/US]; 1121 Tennessee Street #5, San Francisco, CA 94107 (US); LC, Dyoung, Aina, M. [US/US]; 213 Coy Drive, San Jose, CA 95127 (US); YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US); KHAN, Farrah, A. [US/US]; 3617 Central Road #102, Glenview, Illinois 60025 (US)

POLICKY, Jennifer, L. [US/US]; 1511 Jarvis Court, San Jose, CA 95118 (US); AU-YOUNG, Janice [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US); YANG, Junming [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US); BARRAND, Lee [GB/GB]; 18 Chaucer Court, New Dover Road, Canterbury Kent CT1 3AU (GB); WALSH, Roderick, T. [IE/GB]; 8 Boundary Court, St. Laurence Road, Canterbury Kent CT1 3EZ (GB); LO, Terence, P. [CA/US]; 145 Beach Park Boulevard, Apt. 115, Foster City, CA 94041-1941 (US); BOROWSKY, Mark, L. [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US).

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

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, 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, LU, LV, MA, MD, MG, MK, MN, MW, MX, MY, NZ, 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 (GH, GM, KE, LS, MW, MZ, SD, SI, 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, PL, SE, TR); OAPI patent (BF, BI, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, NG, SN, TD, TG).

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 01/66742 A2

(54) Title: G-PROTEIN COUPLED RECEPTORS

(57) Abstract: The invention provides human G-protein coupled receptors (GCREC) and polynucleotides which identify and encode GCREC. 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 GCREC.

WO 01/66742

PCT/US01/06814

G-PROTEIN COUPLED RECEPTORS**TECHNICAL FIELD**

This invention relates to nucleic acid and amino acid sequences of G-protein coupled receptors and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of G-protein coupled receptors.

BACKGROUND OF THE INVENTION

Signal transduction is the general process by which cells respond to extracellular signals. Signal transduction across the plasma membrane begins with the binding of a signal molecule, e.g., a hormone, neurotransmitter, or growth factor, to a cell membrane receptor. The receptor, thus activated, triggers an intracellular biochemical cascade that ends with the activation of an intracellular target molecule, such as a transcription factor. This process of signal transduction regulates all types of cell functions including cell proliferation, differentiation, and gene transcription. The G-protein coupled receptors (GPCRs), encoded by one of the largest families of genes yet identified, play a central role in the transduction of extracellular signals across the plasma membrane. GPCRs have a proven history of being successful therapeutic targets.

GPCRs are integral membrane proteins characterized by the presence of seven hydrophobic transmembrane domains which together form a bundle of antiparallel alpha (α) helices. GPCRs range in size from under 400 to over 1000 amino acids (Strosberg, A.D. (1991) *Eur. J. Biochem.* 196:1-10; Coughlin, S.R. (1994) *Curr. Opin. Cell Biol.* 6:191-197). The amino-terminus of a GPCR is extracellular, is of variable length, and is often glycosylated. The carboxy-terminus is cytoplasmic and generally phosphorylated. Extracellular loops alternate with intracellular loops and link the transmembrane domains. Cysteine disulfide bridges linking the second and third extracellular loops may interact with agonists and antagonists. The most conserved domains of GPCRs are the transmembrane domains and the first two cytoplasmic loops. The transmembrane domains account, in part, for structural and functional features of the receptor. In most cases, the bundle of α helices forms a ligand-binding pocket. The extracellular N-terminal segment, or one or more of the three extracellular loops, may also participate in ligand binding. Ligand binding activates the receptor by inducing a conformational change in intracellular portions of the receptor. In turn, the large, third intracellular loop of the activated receptor interacts with a heterotrimeric guanine nucleotide binding (G) protein complex which mediates further intracellular signaling activities, including the activation of second

WO 01/66742

PCT/US01/06814

messengers such as cyclic AMP (cAMP), phospholipase C, and inositol triphosphate, and the interaction of the activated GPCR with ion channel proteins. (See, e.g., Watson, S. and S. Arkinfall (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 2-6; Bolander, F.F. (1994) Molecular Endocrinology, Academic Press, San Diego CA, pp. 162-176;

5 Baldwin, J.M. (1994) *Curr. Opin. Cell Biol.* 6:180-190.)

GPCRs include receptors for sensory signal mediators (e.g., light and olfactory stimulatory molecules); adenosine, γ -aminobutyric acid (GABA), hepatocyte growth factor, melanocortins, neuropeptide Y, opioid peptides, opsins, somatostatin, tachykinins, vasoactive intestinal polypeptide family, and vasopressin; biogenic amines (e.g., dopamine, epinephrine and norepinephrine, histamine, 10 glutamate (metabotropic effect), acetylcholine (muscarinic effect), and serotonin); chemokines; lipid mediators of inflammation (e.g., prostaglandins and prostanoids, platelet activating factor, and leukotrienes); and peptide hormones (e.g., bombesin, bradykinin, calcitonin, C5a anaphylatoxin, endothelin, follicle-stimulating hormone (FSH), gonadotropin-releasing hormone (GnRH), neurokinin, and thyrotropin-releasing hormone (TRH), and oxytocin). GPCRs which act as receptors for stimuli 15 that have yet to be identified are known as orphan receptors.

The diversity of the GPCR family is further increased by alternative splicing. Many GPCR genes contain introns, and there are currently over 30 such receptors for which splice variants have been identified. The largest number of variations are at the protein C-terminus. N-terminal and cytoplasmic loop variants are also frequent, while variants in the extracellular loops or transmembrane domains are 20 less common. Some receptors have more than one site at which variance can occur. The splicing variants appear to be functionally distinct, based upon observed differences in distribution, signaling, coupling, regulation, and ligand binding profiles (Kilpatrick, G.J. et al. (1999) *Trends Pharmacol. Sci.* 20:294-301).

GPCRs can be divided into three major subfamilies: the rhodopsin-like, secretin-like, and 25 metabotropic glutamate receptor subfamilies. Members of these GPCR subfamilies share similar functions and the characteristic seven transmembrane structure, but have divergent amino acid sequences. The largest family consists of the rhodopsin-like GPCRs, which transmit diverse extracellular signals including hormones, neurotransmitters, and light. Rhodopsin is a photosensitive GPCR found in animal retinas. In vertebrates, rhodopsin molecules are embedded in membranous 30 stacks found in photoreceptor (rod) cells. Each rhodopsin molecule responds to a photon of light by triggering a decrease in cGMP levels which leads to the closure of plasma membrane sodium channels. In this manner, a visual signal is converted to a neural impulse. Other rhodopsin-like GPCRs are directly involved in responding to neurotransmitters. These GPCRs include the receptors for adrenaline (adrenergic receptors), acetylcholine (muscarinic receptors), adenosine, galanin, and glutamate (N-

WO 01/66742

PCT/US01/06814

methyl-D-aspartate/NMDA receptors). (Reviewed in Watson, S. and S. Arkinstall (1994) The G-Protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 7-9, 19-22, 32-35, 130-131, 214-216, 221-222; Habert-Ortoli, E. et al. (1994) Proc. Natl. Acad. Sci. USA 91:9780-9783.)

The galanin receptors mediate the activity of the neuroendocrine peptide galanin, which
5 inhibits secretion of insulin, acetylcholine, serotonin and noradrenaline, and stimulates prolactin and growth hormone release. Galanin receptors are involved in feeding disorders, pain, depression, and Alzheimer's disease (Kask, K. et al. (1997) *Life Sci.* 60:1523-1533). Other nervous system rhodopsin-like GPCRs include a growing family of receptors for lysophosphatidic acid and other
10 lysophospholipids, which appear to have roles in development and neuropathology (Chun, J. et al. (1999) *Cell Biochem. Biophys.* 30:213-242).

The largest subfamily of GPCRs, the olfactory receptors, are also members of the rhodopsin-like GPCR family. These receptors function by transducing odorant signals. Numerous distinct olfactory receptors are required to distinguish different odors. Each olfactory sensory neuron expresses only one type of olfactory receptor, and distinct spatial zones of neurons expressing distinct receptors
15 are found in nasal passages. For example, the RA1c receptor which was isolated from a rat brain library, has been shown to be limited in expression to very distinct regions of the brain and a defined zone of the olfactory epithelium (Raming, K. et al. (1998) *Receptors Channels* 6:141-151). However, the expression of olfactory-like receptors is not confined to olfactory tissues. For example, three rat genes encoding olfactory-like receptors having typical GPCR characteristics showed expression
20 patterns not only in taste and olfactory tissue, but also in male reproductive tissue (Thomas, M.B. et al. (1996) *Gene* 178:1-5).

Members of the secretin-like GPCR subfamily have as their ligands peptide hormones such as secretin, calcitonin, glucagon, growth hormone-releasing hormone, parathyroid hormone, and vasoactive intestinal peptide. For example, the secretin receptor responds to secretin, a peptide
25 hormone that stimulates the secretion of enzymes and ions in the pancreas and small intestine (Watson, supra, pp. 278-283). Secretin receptors are about 450 amino acids in length and are found in the plasma membrane of gastrointestinal cells. Binding of secretin to its receptor stimulates the production of cAMP.

Examples of secretin-like GPCRs implicated in inflammation and the immune response
30 include the EGF module-containing, mucin-like hormone receptor (Emr1) and CD97 receptor proteins. These GPCRs are members of the recently characterized EGF-TM7 receptors subfamily. These seven transmembrane hormone receptors exist as heterodimers in vivo and contain between three and seven potential calcium-binding EGF-like motifs. CD97 is predominantly expressed in leukocytes and is markedly upregulated on activated B and T cells (McKnight, A.J. and S. Gordon

WO 01/66742

PCT/US01/06814

(1998) *J. Leukoc. Biol.* 63:271-280).

The third GPCR subfamily is the metabotropic glutamate receptor family. Glutamate is the major excitatory neurotransmitter in the central nervous system. The metabotropic glutamate receptors modulate the activity of intracellular effectors, and are involved in long-term potentiation (Watson, *supra*, p.130). The Ca^{2+} -sensing receptor, which senses changes in the extracellular concentration of calcium ions, has a large extracellular domain including clusters of acidic amino acids which may be involved in calcium binding. The metabotropic glutamate receptor family also includes pheromone receptors, the GABA_A receptors, and the taste receptors.

Other subfamilies of GPCRs include two groups of chemoreceptor genes found in the nonmammals *Caenorhabditis elegans* and *Caenorhabditis briggsae*, which are distantly related to the mammalian olfactory receptor genes. The yeast pheromone receptors STE2 and STE3, involved in the response to mating factors on the cell membrane, have their own seven-transmembrane signature, as do the cAMP receptors from the slime mold *Dictyostelium discoideum*, which are thought to regulate the aggregation of individual cells and control the expression of numerous developmentally-regulated genes.

GPCR mutations, which may cause loss of function or constitutive activation, have been associated with numerous human diseases (Coughlin, *supra*). For instance, retinitis pigmentosa may arise from mutations in the rhodopsin gene. Furthermore, somatic activating mutations in the thyrotropin receptor have been reported to cause hyperfunctioning thyroid adenomas, suggesting that certain GPCRs susceptible to constitutive activation may behave as protooncogenes (Parma, J. et al. (1993) *Nature* 365:649-651). GPCR receptors for the following ligands also contain mutations associated with human disease: luteinizing hormone (precocious puberty); vasopressin V₂ (X-linked nephrogenic diabetes); glucagon (diabetes and hypertension); calcium (hyperparathyroidism, hypocalcemia, hypercalcaemia); parathyroid hormone (short limbed dwarfism); β_2 -adrenoceptor (obesity, non-insulin-dependent diabetes mellitus); growth hormone releasing hormone (dwarfism); and adrenocorticotropic (glucocorticoid deficiency) (Wilson, S. et al. (1998) *Br. J. Pharmacol.* 125:1387-1392; Stadel, J.M. et al. (1997) *Trends Pharmacol. Sci.* 18:430-437). GPCRs are also involved in depression, schizophrenia, sleeplessness, hypertension, anxiety, stress, renal failure, and several cardiovascular disorders (Horn, F. and G. Vriend (1998) *J. Mol. Med.* 76:464-468).

In addition, within the past 20 years several hundred new drugs have been recognized that are directed towards activating or inhibiting GPCRs. The therapeutic targets of these drugs span a wide range of diseases and disorders, including cardiovascular, gastrointestinal, and central nervous system disorders as well as cancer, osteoporosis and endometriosis (Wilson, *supra*; Stadel, *supra*). For example, the dopamine agonist L-dopa is used to treat Parkinson's disease, while a dopamine antagonist is used to treat schizophrenia and the early stages of Huntington's disease. Agonists and antagonists of

WO 01/66742

PCT/US01/06814

adrenoceptors have been used for the treatment of asthma, high blood pressure, other cardiovascular disorders, and anxiety; muscarinic agonists are used in the treatment of glaucoma and tachycardia; serotonin 5HT_{1D} antagonists are used against migraine; and histamine H₁ antagonists are used against allergic and anaphylactic reactions, hay fever, itching, and motion sickness (Horn, *supra*).

- 5 Recent research suggests potential future therapeutic uses for GPCRs in the treatment of metabolic disorders including diabetes, obesity, and osteoporosis. For example, mutant V₂ vasopressin receptors causing nephrogenic diabetes could be functionally rescued *in vitro* by co-expression of a C-terminal V₂ receptor peptide spanning the region containing the mutations. This result suggests a possible novel strategy for disease treatment (Schönberg, T. et al. (1996) *EMBO J.* 15:1283-1291).
- 10 Mutations in melanocortin-4 receptor (MC4R) are implicated in human weight regulation and obesity. As with the vasopressin V₂ receptor mutants, these MC4R mutants are defective in trafficking to the plasma membrane (Ho, G. and R. G. MacKenzie (1999) *J. Biol. Chem.* 274:35816-35822), and thus might be treated with a similar strategy. The type I receptor for parathyroid hormone (PTH) is a GPCR that mediates the PTH-dependent regulation of calcium homeostasis in the bloodstream. Study
- 15 of PTH/receptor interactions may enable the development of novel PTH receptor ligands for the treatment of osteoporosis (Manstadt, M. et al. (1999) *Am. J. Physiol.* 277:F665-F675).

- The chemokine receptor group of GPCRs have potential therapeutic utility in inflammation and infectious disease. (For review, see Locati, M. and P.M. Murphy (1999) *Annu. Rev. Med.* 50:425-440.) Chemokines are small polypeptides that act as intracellular signals in the regulation of leukocyte
- 20 trafficking, hematopoiesis, and angiogenesis. Targeted disruption of various chemokine receptors in mice indicates that these receptors play roles in pathologic inflammation and in autoimmune disorders such as multiple sclerosis. Chemokine receptors are also exploited by infectious agents, including herpesviruses and the human immunodeficiency virus (HIV-1) to facilitate infection. A truncated version of chemokine receptor CCR5, which acts as a coreceptor for infection of T-cells by HIV-1,
- 25 results in resistance to AIDS, suggesting that CCR5 antagonists could be useful in preventing the development of AIDS.

- The discovery of new G-protein coupled receptors 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, neurological, cardiovascular, gastrointestinal,
- 30 autoimmune/inflammatory, and metabolic disorders, and viral infections, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of G-protein coupled receptors.

SUMMARY OF THE INVENTION

WO 01/66742

PCT/US01/06814

The invention features purified polypeptides, G-protein coupled receptors, referred to collectively as "GCREC" and individually as "GCREC-1," "GCREC-2," "GCREC-3," "GCREC-4," "GCREC-5," "GCREC-6," "GCREC-7," "GCREC-8," "GCREC-9," "GCREC-10," "GCREC-11," "GCREC-12," "GCREC-13," "GCREC-14," "GCREC-15," "GCREC-16," "GCREC-17," "GCREC-18," "GCREC-19," "GCREC-20," and "GCREC-21." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-21.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-21. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:22-42.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90%

WO 01/66742

PCT/US01/06814

sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) culturing a cell under conditions suitable for
5 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 comprising an amino acid sequence selected from the group consisting of a) an amino acid
10 sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence
15 selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e)
20 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 comprising a polynucleotide sequence
25 selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising
30 at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

WO 01/66742

PCT/US01/06814

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to 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 comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional GCREC, 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 comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as

WO 01/66742

PCT/US01/06814

an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional GCREC, 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 comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the

WO 01/66742

PCT/US01/06814

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:22-42, 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 comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to 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 comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to 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.

BRIEF DESCRIPTION OF THE TABLES

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

Table 2 shows the GenBank identification number and annotation of the nearest GenBank 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

WO 01/66742

PCT/US01/06814

motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide
5 sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and
10 polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these
15 may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a
20 reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although
25 any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is
30 not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"GCREC" refers to the amino acid sequences of substantially purified GCREC 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.

WO 01/66742

PCT/US01/06814

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of GCREC. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of GCREC either by directly interacting with GCREC or by acting on components of the biological pathway in which GCREC participates.

5 An "allelic variant" is an alternative form of the gene encoding GCREC. 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.

10 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 GCREC include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as GCREC or a polypeptide with at least one functional characteristic of GCREC. Included within this definition are

15 polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding GCREC, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding GCREC. 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 GCREC.

20 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 GCREC 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

25 values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

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

30 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

WO 01/66742

PCT/US01/06814

in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of GCREC. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of GCREC either by
5 directly interacting with GCREC or by acting on components of the biological pathway in which GCREC 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 GCREC polypeptides can be prepared using intact polypeptides or using
10 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
15 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
20 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
25 phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring
30 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"

WO 01/66742

PCT/US01/06814

refers to the capability of the natural, recombinant, or synthetic GCREC, 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 GCREC or fragments of GCREC 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 Pirap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

“Conservative amino acid substitutions” are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val

WO 01/66742

PCT/US01/06814

	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
5	Ser	Cys, Trp
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

10

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.

15 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.

20 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.

25 A "fragment" is a unique portion of GCRC or the polynucleotide encoding GCRC 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 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 30 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 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 35 by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:22-42 comprises a region of unique polynucleotide sequence that

WO 01/66742

PCT/US01/06814

specifically identifies SEQ ID NO:22-42, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:22-42 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:22-42 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:22-42 and the region of SEQ ID NO:22-42 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-21 is encoded by a fragment of SEQ ID NO:22-42. A fragment of SEQ ID NO:1-21 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-21. For example, a fragment of SEQ ID NO:1-21 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-21. The precise length of a fragment of SEQ ID NO:1-21 and the region of SEQ ID NO:1-21 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MBGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search

WO 01/66742

PCT/US01/06814

Tool (BLAST) (Altschul, S.F. et al. (1990) *J. Mol. Biol.* 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn" that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/b2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

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,

WO 01/66742

PCT/US01/06814

explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

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

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

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.

WO 01/66742

PCT/US01/06814

Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is 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., C_{50} or R_{50} 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).

WO 01/66742

PCT/US01/06814

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 GCREC 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 GCREC which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

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

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an GCREC 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

WO 01/66742

PCT/US01/06814

cell type depending on the enzymatic milieu of GCREC.

"Probe" refers to nucleic acid sequences encoding GCREC, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols. A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such 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 PrimoOU 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

WO 01/66742

PCT/US01/06814

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 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 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 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.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

WO 01/66742

PCT/US01/06814

The term "sample" is used in its broadest sense. A sample suspected of containing GCREC, nucleic acids encoding GCREC, or fragments thereof may comprise a bodily fluid, an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

5 The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope
10 A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which
15 they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers,
20 microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" 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
25 cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells"
30 includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic

WO 01/66742

PCT/US01/06814

acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in
5 vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention
10 into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999)
15 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 "allele" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant
20 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
25 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
30 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

WO 01/66742

PCT/US01/06814

certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human G-protein coupled receptors (GCREC),
 5 the polynucleotides encoding GCREC, and the use of these compositions for the diagnosis, treatment, or
 prevention of cell proliferative, neurological, cardiovascular, gastrointestinal,
 autoimmune/inflammatory, and metabolic disorders, and viral infections.

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
 10 single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted
 by both a polypeptide sequence identification number (Polypeptide SEQ ID NO.) and an Incyte
 polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is
 denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO.) and an
 Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

15 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.
 20 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
 25 polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3
 shows the number of amino acid residues in each polypeptide. Column 4 shows potential
 phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS
 program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI).
 Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7
 30 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 G-protein coupled receptors. For example, SEQ
 ID NO:7 is 85% identical, from residue M1 to residue V306, to murine odorant receptor MOR83

WO 01/66742

PCT/US01/06814

(GenBank ID g6178006) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 5.5e-141, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:7 also contains a 7 transmembrane receptor (rhodopsin family) domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:7 is an olfactory G-protein coupled receptor. SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-21 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were 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 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:22-42 or that distinguish between SEQ ID NO:22-42 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA 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 genomic sequences in column 5 relative to their respective full length sequences.

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, 6871486H1 is the identification number of an Incyte cDNA sequence, and BRAGNON02 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 70171099V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g5743982) which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by GenScan analysis of genomic DNA. For example, GNN.g6671985_006 is the

WO 01/66742

PCT/US01/06814

identification number of a Genscan-predicted coding sequence, with g6671955 being the GenBank identification number of the sequence to which Genscan was applied. 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. For example, U12289894_00001 represents a "stitched" sequence in which 2289894 is the identification number of the cluster of sequences to which the algorithm was applied, and 00001 is the number of the prediction generated by the 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 GCREC variants. A preferred GCREC 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 GCREC amino acid sequence, and which contains at least one functional or structural characteristic of GCREC.

The invention also encompasses polynucleotides which encode GCREC. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:22-42, which encodes GCREC. The polynucleotide sequences of SEQ ID NO:22-42, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding GCREC. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding GCREC. 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:22-42 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:22-42.

WO 01/66742

PCT/US01/06814

Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of GCREC.

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 GCREC, 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 polynucleotide sequence of naturally occurring GCREC, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode GCREC and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring GCREC under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding GCREC or 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 GCREC and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode GCREC and GCREC derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding GCREC or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:22-42 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Bregar (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or

WO 01/66742

PCT/US01/06814

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 GCRC 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.

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

WO 01/66742

PCT/US01/06814

does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

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

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter GCREC-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULAR BREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of GCREC, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random

WO 01/66742

PCT/US01/06814

point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding GCREC may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Carothers, M.H. et al. (1980) *Nucleic Acids Symp. Ser.* 7:215-223; and Hohn, T. et al. (1980) *Nucleic Acids Symp. Ser.* 7:225-232.) Alternatively, GCREC itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) *Proteins, Structures and Molecular Properties*, WH Freeman, New York NY, pp. 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 GCREC, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, *supra*, pp. 28-53.)

In order to express a biologically active GCREC, the nucleotide sequences encoding GCREC or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding GCREC. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding GCREC. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding GCREC and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may 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

WO 01/66742

PCT/US01/06814

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.)

5 Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding GCREC 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.)

10 A variety of expression vector/host systems may be utilized to contain and express sequences encoding GCREC. 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 Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and 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.

15 In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding GCREC. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding GCREC can be achieved using a multifunctional *E. coli* vector such as PBLUE:SCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding GCREC into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro

WO 01/66742

PCT/US01/06814

transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Hecke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of GCREC are needed, e.g. for the production of antibodies, vectors which direct high level expression of GCREC may be used. For example, vectors containing
5 the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of GCREC. 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
10 of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, *supra*; 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 GCREC. Transcription of sequences encoding GCREC may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used
15 alone or in combination with the omega leader sequence from TMV (Takanatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated
20 transfection. (See, e.g., *The McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding GCREC may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader
25 sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses GCREC in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of
30 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

WO 01/66742

PCT/US01/06814

GCREC in cell lines is preferred. For example, sequences encoding GCREC 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
5 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,
10 but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *aprt* 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
15 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
20 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. (1993) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is
25 also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding GCREC is inserted within a marker gene sequence, transformed cells containing sequences encoding GCREC can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding GCREC under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates
30 expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding GCREC and that express GCREC 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

WO 01/66742

PCT/US01/06814

chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of GCREC 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

5 activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on GCREC is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Jampton, 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

10 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 GCREC include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide.

15 Alternatively, the sequences encoding GCREC, 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

20 (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 GCREC may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein

25 produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode GCREC may be designed to contain signal sequences which direct secretion of GCREC through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the

30 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

WO 01/66742

PCT/US01/06814

(e.g., CHO, HeLa, MDCK, IIEK293, 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 GCREC 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 GCREC protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of GCREC 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 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 immunofluorescence purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the GCREC encoding sequence and the heterologous protein sequence, so that GCREC 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.

In a further embodiment of the invention, synthesis of radiolabeled GCREC 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 precursor, for example, ³⁵S-methionine.

GCREC of the present invention or fragments thereof may be used to screen for compounds that specifically bind to GCREC. At least one and up to a plurality of test compounds may be screened for specific binding to GCREC. 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 GCREC, 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 GCREC binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the

WO 01/66742

PCT/US01/06814

compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express GCREC, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing GCREC or cell membrane fractions which contain GCREC are then contacted with a test compound and binding, stimulation, or inhibition of activity of either GCREC 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 GCREC, either in solution or affixed to a solid support, and detecting the binding of GCREC 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.

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

In another embodiment, polynucleotides encoding GCREC 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

WO 01/66742

PCT/US01/06814

system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marin, J.D. (1996) *Clin. Invest.* 97:1999-2002; Wagner, K.U. et al. (1997) *Nucleic Acids Res.* 25:4323-4330). Transfected 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
5 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 GCREC 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
10 into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) *Science* 282:1145-1147).

Polynucleotides encoding GCREC 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 GCREC is injected into animal ES cells, and the injected sequence
15 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 GCREC, e.g., by secreting GCREC in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

20 THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of GCREC and G-protein coupled receptors. In addition, the expression of GCREC is closely associated with ovarian tumor, prostate, white blood cells, cerebellar, and brain tissues. Therefore, GCREC appears to play a role in cell proliferative, neurological, cardiovascular,
25 gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections. In the treatment of disorders associated with increased GCREC expression or activity, it is desirable to decrease the expression or activity of GCREC. In the treatment of disorders associated with decreased GCREC expression or activity, it is desirable to increase the expression or activity of GCREC.

30 Therefore, in one embodiment, GCREC 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 GCREC. 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), myotofibrosis, paroxysmal nocturnal hemoglobinuria,

WO 01/66742

PCT/US01/06814

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,

5 salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral acoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial

10 and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental

15 retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders,

20 seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebotrombosis, vascular

25 tumors, complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic

30 endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina,

35 pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer,

WO 01/66742

PCT/US01/06814

cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha₁-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a metabolic disorder such as diabetes, obesity, and osteoporosis; and an infection by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and togavirus.

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

In a further embodiment, a composition comprising a substantially purified GCREC in 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 GCREC including, but not limited to, those provided above.

WO 01/66742

PCT/US01/06814

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

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

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

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate 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.

An antagonist of GCREC may be produced using methods which are generally known in the art. In particular, purified GCREC may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind GCREC. Antibodies to GCREC 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.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with GCREC or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysollecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and diitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guérin) and Corynebacterium parvum are especially preferable.

WO 01/66742

PCT/US01/06814

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to GCREC 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 GCREC 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 GCREC 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 GCREC-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotype 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 GCREC 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. et al. (1989) *Science* 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such

WO 01/66742

PCT/US01/06814

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

5 Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for GCREC. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of GCREC-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
10 GCREC epitopes, represents the average affinity, or avidity, of the antibodies for GCREC. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular GCREC 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 GCREC-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a
15 ranging from about 10^5 to 10^7 L/mole are preferred for use in immunoprecipitation and similar procedures which ultimately require dissociation of GCREC, preferably in active form, from the antibody (Catty, D. (1988) *Antibodies, Volume I: A Practical Approach*, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) *A Practical Guide to Monoclonal Antibodies*, John Wiley & Sons, New York NY).

20 The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of GCREC-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for
25 antibody quality and usage in various applications, are generally available. (See, e.g., Catty, *supra*, and Coligan et al. *supra*.)

In another embodiment of the invention, the polynucleotides encoding GCREC, 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, RNA,
30 PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding GCREC. 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 GCREC. (See, e.g., Agrawal, S., ed. (1996) *Antisense Therapeutics*, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense

WO 01/66742

PCT/US01/06814

sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding GCREC may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Sonni (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA.* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in GCREC expression or regulation causes disease, the expression of GCREC 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 GCREC are treated by constructing mammalian expression vectors encoding GCREC and introducing these vectors by mechanical means into GCREC-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic

WO 01/66742

PCT/US01/06814

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).

5 Expression vectors that may be effective for the expression of GCREC 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). GCREC may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus
10 (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 dexamethasone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the
15 FK506/tapamycin 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 GCREC 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
20 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.

25 In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to GCREC expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding GCREC under the control of an independent promoter or the retrovirus long 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
30 required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, J. 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.

WO 01/66742

PCT/US01/06814

(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 GCREC to cells which have one or more genetic abnormalities with respect to the expression of GCREC. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) *Annu. Rev. Nutr.* 19:511-544 and Verma, I.M. and N. Senita (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 GCREC to target cells which have one or more genetic abnormalities with respect to the expression of GCREC. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing GCREC to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 69:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent 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 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.*

WO 01/66742

PCT/US01/06814

Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding GCREC 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, 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 GCREC into the alphavirus genome in place of the capsid-coding region results in the production of a large number of GCREC-coding RNAs and the synthesis of high levels of GCREC in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of GCREC into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can 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., Gea, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, *Molecular and Immunologic Approaches*, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozymic action involves sequence-specific hybridization of the ribozyme

WO 01/66742

PCT/US01/06814

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 GCREC.

5 Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary
10 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 sequences
15 encoding GCREC. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible
20 modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine,
25 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 GCREC. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming
30 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 GCREC expression or activity, a compound which specifically inhibits expression of the

WO 01/66742

PCT/US01/06814

polynucleotide encoding GCREC may be therapeutically useful, and in the treatment of disorders associated with decreased GCREC expression or activity, a compound which specifically promotes expression of the polynucleotide encoding GCREC may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding GCREC 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 GCREC are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding GCREC. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (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).

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. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such

WO 01/66742

PCT/US01/06814

therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient.

5 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 GCREC, antibodies to GCREC, and mimetics, agonists, antagonists, or inhibitors of GCREC.

The compositions utilized in this invention may be administered by any number of routes
10 including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, ocular, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form.

These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case
15 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 needle
20 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 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
25 macromolecules comprising GCREC or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, GCREC 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
30 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

WO 01/66742

PCT/US01/06814

administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example GCREC or fragments thereof, antibodies of GCREC, and agonists, antagonists or inhibitors of GCREC, 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 ED₅₀ (the dose therapeutically effective in 50% of the population) or LD₅₀ (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 LD₅₀/ED₅₀ 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 ED₅₀ 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 μ g to 100,000 μ g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind GCREC may be used for the diagnosis of disorders characterized by expression of GCREC, or in assays to monitor patients being treated with GCREC or agonists, antagonists, or inhibitors of GCREC. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for GCREC include methods which utilize the antibody and a label to detect GCREC 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

WO 01/66742

PCT/US01/06814

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 GCREC, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of GCREC expression. Normal or standard values for GCREC expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to GCREC under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of GCREC 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 GCREC 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 GCREC may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of GCREC, and to monitor regulation of GCREC levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding GCREC or closely related molecules may be used to identify nucleic acid sequences which encode GCREC. 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 probe identifies only naturally occurring sequences encoding GCREC, 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 GCREC encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:22-42 or from genomic sequences including promoters, enhancers, and introns of the GCREC gene.

Means for producing specific hybridization probes for DNAs encoding GCREC include the cloning of polynucleotide sequences encoding GCREC or GCREC 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 ^{32}P or ^{35}S , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

WO 01/66742

PCT/US01/06814

Polynucleotide sequences encoding GCREC may be used for the diagnosis of disorders associated with expression of GCREC. 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

5 hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological

10 disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess,

15 suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central

20 nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia,

25 catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebotrombosis, vascular tumors, complications of thrombolysis, balloon

30 angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus

35 erythematosus, carcinoma heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart

WO 01/66742

PCT/US01/06814

disease, congenital heart disease, and complications of cardiac transplantation; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, aural or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal

5 obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic

10 obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha-1-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of

15 pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemias, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact

20 dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxicity, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypercosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis,

25 psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a metabolic disorder such as diabetes, obesity, and osteoporosis; and an infection by a viral agent

30 classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and togavirus. The polynucleotide sequences encoding GCRC 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 microarrays utilizing

35 fluids or tissues from patients to detect altered GCRC expression. Such qualitative or quantitative

WO 01/66742

PCT/US01/06814

methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding GCREC may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding GCREC may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding GCREC in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of GCREC, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding GCREC, 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 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 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 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 GCREC 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

WO 01/66742

PCT/US01/06814

GCREC, or a fragment of a polynucleotide complementary to the polynucleotide encoding GCREC, 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.

5 In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding GCREC 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
10 derived from the polynucleotide sequences encoding GCREC 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
15 fluorescently labeled, which allows detection of the amplicons 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
20 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 GCREC include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from
25 standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Dupliss, 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 interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

30 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

WO 01/66742

PCT/US01/06814

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, GCREC, fragments of GCREC, or antibodies specific for GCREC 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.

10 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 Selthamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number
15 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
20 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.

25 Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) *Mol. Carcinog.* 24:153-159; Steiner, S. and N.L. Anderson (2000)
30 *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

WO 01/66742

PCT/US01/06814

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
5 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.

10 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
15 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
20 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
25 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
30 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

WO 01/66742

PCT/US01/06814

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 GCREC to quantify the levels of GCREC 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; Mendozá, 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. Scilheimer (1997) *Electrophoresis* 18:533-537), so proteomic 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 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., Breiman, 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

WO 01/66742

PCT/US01/06814

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 GCREC 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.I. et al. (1997) *Nat. Genet.* 15:345-355; Price, C.M. (1993) *Blood Rev.* 7:127-134; and Trask, B.J. (1991) *Trends Genet.* 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) *Proc. Natl. Acad. Sci. USA* 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding GCREC 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, GCREC, its catalytic or immunogenic fragments, or

WO 01/66742

PCT/US01/06814

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 GCREC and the agent being tested may be measured.

5 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 GCREC, or fragments thereof, and washed. Bound GCREC is then detected by methods well known in the art. Purified GCREC can also be coated directly onto plates for use in the aforementioned drug screening techniques.
10 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 GCREC specifically compete with a test compound for binding GCREC.
15 In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with GCREC.

In additional embodiments, the nucleotide sequences which encode GCREC may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such
20 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.

25 The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/186,854, U.S. Ser. No. 60/188,384, U.S. Ser. No. 60/190,453, U.S. Ser. No. 60/190,730 are hereby expressly incorporated by reference.

EXAMPLES

30 I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESTEQ GOLD database (Incyte 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

WO 01/66742

PCT/US01/06814

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
 5 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).

10 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 SUPERSCRIPT 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
 15 oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g.,
 20 PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBE-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.

25 II. Isolation of cDNA Clones

Plasmids obtained as described in Example 1 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
 30 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

WO 01/66742

PCT/US01/06814

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).

5 **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
10 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
15 PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software, or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

20 The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS,
25 DOMO, PRODOM, and hidden Markov model (HMM)-based protein family 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
30 cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Gauscan-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

WO 01/66742

PCT/US01/06814

polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, 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, 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 multiple sequence 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 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 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:22-42. 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 G-protein coupled receptors were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpt1 and gbhg). 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 G-protein coupled receptors, the encoded polypeptides were analyzed by querying against PFAM models for G-protein coupled receptors. Potential G-protein coupled receptors were also identified by homology to Incyte cDNA sequences that had been annotated as G-protein coupled

WO 01/66742

PCT/US01/06814

receptors. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbprl public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on 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 gbprl 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

WO 01/66742

PCT/US01/06814

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 GCRFC Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:22-42 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:22-42 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Génethon 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, or 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 Génethon 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 "GenMap'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:23 was mapped to chromosome 16 within the interval from 57.8 to 71.4 centiMorgans.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene

WO 01/66742

PCT/US01/06814

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 LITESEQ (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:

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

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 GCRC 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,

WO 01/66742

PCT/US01/06814

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 GCREC. cDNA sequences and cDNA library/tissue information are found in the JFESSEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of GCREC Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one 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 Mg²⁺, (NH₄)₂SO₄, 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 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl 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 µl to 10 µl 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,

WO 01/66742

PCT/US01/06814

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
5 religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

10 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 recoveries were reamplified
15 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
20 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:22-42 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is
25 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
30 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.

WO 01/66742

PCT/US01/06814

membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

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., Baldeschweller, *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 (Schena (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., Schena, 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.

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

WO 01/66742

PCT/US01/06814

transcription reaction is performed in a 25 μ l volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Lucyte). Specific control poly(A)⁺ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and

5 incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 μ l of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and

10 resuspended in 14 μ l 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are

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

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water

20 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

25 Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average concentration of 100 ng/ μ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water.

30 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 μ l of sample mixture consisting of 0.2 μ g each of Cy3 and

35 Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample

WO 01/66742

PCT/US01/06814

mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for
5 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.

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
10 at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate 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
20 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
25 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, 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 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
35 measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping

WO 01/66742

PCT/US01/06814

emission 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 GCRC-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring GCRC. 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 GCRC. 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 GCRC-encoding transcript.

XII. Expression of GCRC

Expression and purification of GCRC is achieved using bacterial or virus-based expression systems. For expression of GCRC 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 GCRC upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of GCRC 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 GCRC 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, GCRC 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 01/66742

PCT/US01/06814

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

XIII. Functional Assays

GCREC function is assessed by expressing the sequences encoding GCREC 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 GCREC on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding GCREC and either CD64 or CD64-GFP. CD64 and

WO 01/66742

PCT/US01/06814

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 GCREC and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of GCREC Specific Antibodies

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

XV. Purification of Naturally Occurring GCREC Using Specific Antibodies

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

XVI. Identification of Molecules Which Interact with GCREC

WO 01/66742

PCT/US01/06814

Molecules which interact with GCREC may include agonists and antagonists, as well as molecules involved in signal transduction, such as G proteins. GCREC, or a fragment thereof, is labeled with ¹²⁵I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) A fragment of GCREC includes, for example, a fragment comprising one or more of the
5 three extracellular loops, the extracellular N-terminal region, or the third intracellular loop. Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled GCREC, washed, and any wells with labeled GCREC complex are assayed. Data obtained using different concentrations of GCREC are used to calculate values for the number, affinity, and association of GCREC with the candidate ligand molecules.

10 Alternatively, molecules interacting with GCREC 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). GCREC 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
15 the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Potential GCREC agonists or antagonists may be tested for activation or inhibition of GCREC receptor activity using the assays described in sections XVII and XVIII. Candidate molecules may be selected from known GPCR agonists or antagonists, peptide libraries, or combinatorial chemical
20 libraries.

Methods for detecting interactions of GCREC with intracellular signal transduction molecules such as G proteins are based on the premise that internal segments or cytoplasmic domains from an orphan G protein-coupled seven transmembrane receptor may be exchanged with the analogous domains of a known G protein-coupled seven transmembrane receptor and used to identify the G-
25 proteins and downstream signaling pathways activated by the orphan receptor domains (Kobilka, B.K. et al. (1988) *Science* 240:1310-1316). In an analogous fashion, domains of the orphan receptor may be cloned as a portion of a fusion protein and used in binding assays to demonstrate interactions with specific G proteins. Studies have shown that the third intracellular loop of G protein-coupled seven transmembrane receptors is important for G protein interaction and signal transduction
30 (Conklin, B.R. et al. (1993) *Cell* 73:631-641). For example, the DNA fragment corresponding to the third intracellular loop of GCREC may be amplified by the polymerase chain reaction (PCR) and subcloned into a fusion vector such as pGEX (Pharmacia Biotech). The construct is transformed into an appropriate bacterial host, induced, and the fusion protein is purified from the cell lysate by glutathione-Sepharose 4B (Pharmacia Biotech) affinity chromatography.

WO 01/66742

PCT/US01/06814

For *in vitro* binding assays, cell extracts containing G proteins are prepared by extraction with 50 mM Tris, pH 7.8, 1 mM EGTA, 5 mM MgCl₂, 20 mM CHAPS, 20% glycerol, 10 µg of both aprotinin and leupeptin, and 20 µl of 50 mM phenylmethylsulfonyl fluoride. The lysate is incubated on ice for 45 min with constant stirring, centrifuged at 23,000 g for 15 min at 4°C, and the supernatant is collected. 750 µg of cell extract is incubated with glutathione S-transferase (GST) fusion protein beads for 2 h at 4°C. The GST beads are washed five times with phosphate-buffered saline. Bound G subunits are detected by [³²P]ADP-ribosylation with pertussis or cholera toxins. The reactions are terminated by the addition of SDS sample buffer (4.6% (w/v) SDS, 10% (v/v) β-mercaptoethanol, 20% (w/v) glycerol, 95.2 mM Tris-HCl, pH 6.8, 0.01% (w/v) bromophenol blue). The [³²P]ADP-labeled proteins are separated on 10% SDS-PAGE gels, and autoradiographed. The separated proteins in these gels are transferred to nitrocellulose paper, blocked with blotto (5% nonfat dried milk, 50 mM Tris-HCl (pH 8.0), 2 mM CaCl₂, 80 mM NaCl, 0.02% Na₂S₂O₃, and 0.2% Nonidet P-40) for 1 hour at room temperature, followed by incubation for 1.5 hours with Gα subtype selective antibodies (1:500; Calbiochem-Novabiochem). After three washes, blots are incubated with horseradish peroxidase (HRP)-conjugated goat anti-rabbit immunoglobulin (1:2000; Cappel, Westchester PA) and visualized by the chemiluminescence-based ECL method (Amersham Corp.).

XVII. Demonstration of GCREC Activity

An assay for GCREC activity measures the expression of GCREC on the cell surface. cDNA encoding GCREC is transfected into an appropriate mammalian cell line. Cell surface proteins are labeled with biotin as described (de la Fuente, M.A. et al. (1997) *Blood* 90:2398-2405). Immunoprecipitations are performed using GCREC-specific antibodies, and immunoprecipitated samples are analyzed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled immunoprecipitant is proportional to the amount of GCREC expressed on the cell surface.

In the alternative, an assay for GCREC activity is based on a prototypical assay for ligand/receptor-mediated modulation of cell proliferation. This assay measures the rate of DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding GCREC is added to quiescent 3T3 cultured cells using transfection methods well known in the art. The transiently transfected cells are then incubated in the presence of [³H]thymidine, a radioactive DNA precursor molecule. Varying amounts of GCREC ligand are then added to the cultured cells. Incorporation of [³H]thymidine into acid-precipitable DNA is measured over an appropriate time interval using a radiolotope counter, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold GCREC ligand concentration range is indicative of receptor activity. One unit of activity per milliliter is defined as the

WO 01/66742

PCT/US01/06814

concentration of GCREC producing a 50% response level, where 100% represents maximal incorporation of [³H]thymidine into acid-precipitable DNA (McKay, I. and I. Leigh, eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York NY, p. 73.)

In a further alternative, the assay for GCREC activity is based upon the ability of GPCR family proteins to modulate G protein-activated second messenger signal transduction pathways (e.g., cAMP; Gaudin, P. et al. (1998) *J. Biol. Chem.* 273:4990-4996). A plasmid encoding full length GCREC is transfected into a mammalian cell line (e.g., Chinese hamster ovary (CHO) or human embryonic kidney (HEK-293) cell lines) using methods well-known in the art. Transfected cells are grown in 12-well trays in culture medium for 48 hours, then the culture medium is discarded, and the attached cells are gently washed with PBS. The cells are then incubated in culture medium with or without ligand for 30 minutes, then the medium is removed and cells lysed by treatment with 1 M perchloric acid. The cAMP levels in the lysate are measured by radioimmunoassay using methods well-known in the art. Changes in the levels of cAMP in the lysate from cells exposed to ligand compared to those without ligand are proportional to the amount of GCREC present in the transfected cells.

To measure changes in inositol phosphate levels, the cells are grown in 24-well plates containing 1×10^5 cells/well and incubated with inositol-free media and [³H]inositol, 2 μ Ci/well, for 48 hr. The culture medium is removed, and the cells washed with buffer containing 10 mM LiCl followed by addition of ligand. The reaction is stopped by addition of perchloric acid. Inositol phosphates are extracted and separated on Dowex AG1-X8 (Bio-Rad) anion exchange resin, and the total labeled inositol phosphates counted by liquid scintillation. Changes in the levels of labeled inositol phosphate from cells exposed to ligand compared to those without ligand are proportional to the amount of GCREC present in the transfected cells.

XVIII. Identification of GCREC Ligands

GCREC is expressed in a eukaryotic cell line such as CHO (Chinese Hamster Ovary) or HEK (Human Embryonic Kidney) 293 which have a good history of GPCR expression and which contain a wide range of G-proteins allowing for functional coupling of the expressed GCREC to downstream effectors. The transformed cells are assayed for activation of the expressed receptors in the presence of candidate ligands. Activity is measured by changes in intracellular second messengers, such as cyclic AMP or Ca²⁺. These may be measured directly using standard methods well known in the art, or by the use of reporter gene assays in which a luminescent protein (e.g. firefly luciferase or green fluorescent protein) is under the transcriptional control of a promoter responsive to the stimulation of protein kinase C by the activated receptor (Milligan, G. et al. (1996) *Trends Pharmacol. Sci.* 17:235-237). Assay technologies are available for both of these second messenger systems to allow high throughput readout.

WO 01/66742

PCT/US01/06814

in multi-well plate format, such as the adenylyl cyclase activation FlashPlate Assay (NEN Life Sciences Products), or fluorescent Ca^{2+} indicators such as Fluo-4 AM (Molecular Probes) in combination with the FLIPR fluorimetric plate reading system (Molecular Devices). In cases where the physiologically relevant second messenger pathway is not known, GCREC may be coexpressed with the G-proteins $G_{\alpha 12/16}$ which have been demonstrated to couple to a wide range of G-proteins (Offermanns, S. and M.L. Simon (1995) *J. Biol. Chem.* 270:15175-15180), in order to funnel the signal transduction of the GCREC through a pathway involving phospholipase C and Ca^{2+} mobilization. Alternatively, GCREC may be expressed in engineered yeast systems which lack endogenous GPCRs, thus providing the advantage of a null background for GCREC activation screening. These yeast systems substitute a human GPCR and G_{α} protein for the corresponding components of the endogenous yeast pheromone receptor pathway. Downstream signaling pathways are also modified so that the normal yeast response to the signal is converted to positive growth on selective media or to reporter gene expression (Broach, J.R. and J. Thorne (1996) *Nature* 384 (supp.):14-16). The receptors are screened against putative ligands including known GPCR ligands and other naturally occurring bioactive molecules. Biological extracts from tissues, biological fluids and cell supernatants are also screened.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide Seq. ID No.	Incyte Polypeptide ID	Polynucleotide Seq. ID No.	Incyte Polynucleotide ID
536462	1	536482C01	22	536482C01
1316020	2	1316020C01	23	1316020C01
2816437	3	2816437C01	24	2816437C01
2289894	4	2289894C01	25	2289894C01
7066050	5	7066050C01	26	7066050C01
5376785	6	5376785C01	27	5376785C01
3082743	7	3082743C01	28	3082743C01
7472445	8	7472445C01	29	7472445C01
7472445	9	7472445C01	30	7472445C01
7472445	10	7472445C01	31	7472445C01
7472434	11	7472434C01	32	7472434C01
7472435	12	7472435C01	33	7472435C01
7472438	13	7472438C01	34	7472438C01
7472439	14	7472439C01	35	7472439C01
7472440	15	7472440C01	36	7472440C01
7472443	16	7472443C01	37	7472443C01
7472445	17	7472445C01	38	7472445C01
7472445	18	7472445C01	39	7472445C01
7472451	19	7472451C01	40	7472451C01
7472456	20	7472456C01	41	7472456C01
7472457	21	7472457C01	42	7472457C01

WO 01/66742

PC/T/US01/06814

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	SeaBank HomeLog
2	1316020	9927203	4.10E-21	alpha 1C adrenergic receptor isoform 2 (Homo sapiens) (Miyasawa, K., et al. (1995) Cloning, functional expression and tissue distribution of human alpha 1C-adrenoceptor splice variants. FEBS Lett. 363 (3), 256-260.)
4	2289894	9992582	3.20E-35	G protein-coupled seven-transmembrane receptor (Oryzias latipes) (Yasuda, K., et al. (1995) Molecular cloning of a fish gene encoding a novel seven-transmembrane receptor related distantly to catecholamine, histamine, and serotonin receptors. Biochim. Biophys. Acta 1235, 467-469.)
5	7066050	96541395	2.30E-175	growth factor-regulated G protein-coupled receptor ARG-1 (Rattus norvegicus) (Glickman, M., et al. (1998) Molecular cloning, tissue-specific expression, and chromosomal localization of a novel nerve growth factor- regulated G-protein-coupled receptor, nrg-1. Mol. Cell. Neurosci. 14, 444-452.)
6	5376785	9460318	5.30E-72	G-protein coupled receptor (Mus musculus) (Harrigan, H.T., et al. (1991) Identification of a gene induced by glucocorticoids in murine T-cells: a potential G protein-coupled receptor. Mol. Endocrinol. 5, 1311-1318.)

WO 01/66742

PC1/US01/06814

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
7	3082743	66178006	5.50E-141	odorant receptor MOR33 (Mus musculus) (Tauboel, A. et al. (1999) Olfactory neurons expressing closely linked and homologous odorant receptor genes tend to project their axons to neighboring glomeruli on the olfactory bulb. J. Neurosci. 19:8409-8418.)
8	7472361	03927808	1.80E-85	olfactory receptor-like protein COR3/beta (Gallus gallus) (Kallman, M. et al. (1993) Primary sequence, evolution, and repetitive elements of the Gallus gallus (chicken) beta-globin cluster. Genomics 18: 616-625.)
9	7472363	66532001	1.90E-90	odorant receptor S19 (Mus musculus)
10	7472364	54680268	2.00E-95	odorant receptor S46 (Mus musculus) (Mahlitz, B. et al. (1999) Combinatorial receptor codes for odors. Cell 96: 713-723.)
11	7472434	65869927	2.20E-98	olfactory receptor (Mus musculus)
12	7472435	65869916	4.00E-138	olfactory receptor (Mus musculus)
13	7472438	65963968	1.00E-141	odorant receptor M72 (Mus musculus) (Zheng, C. et al. (2000) Peripheral olfactory projections are differentially affected in mice deficient in a cyclic nucleotide-gated channel subunit. Neuron 26: 81-91.)

WO 01/66742

PCT/US01/06814

Table 2 (cont.)

Polypeptide Seq ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
14	7472439	g11692541	1.00E-123	odorant receptor K23 [Mus musculus] (Xie, S. Y. et al. (2000) Characterization of a cluster comprising approximately 100 odorant receptor genes in mouse. Mamm. Genome 11: 1070-1078.)
15	7472440	g11692535	1.00E-135	odorant receptor K21 [Mus musculus] (Xie, S. Y. et al. (2000) Characterization of a cluster comprising approximately 100 odorant receptor genes in mouse. Mamm. Genome 11: 1070-1078.)
16	7472443	g1246534	3.50E-83	olfactory receptor 4 [Gallus gallus] (Leibowitz, M. et al. (1996) Avian olfactory receptors: differentiation of olfactory neurons under normal and experimental conditions. Dev. Biol. 175:118-131.)
17	7472445	g5453092	1.00E-102	olfactory receptor [Mus musculus domesticus] (Fouquier, S. et al. (2000) The olfactory receptor gene repertoire in primates and mouse: evidence for reduction of the functional fraction in primates. Proc. Natl. Acad. Sci. U.S.A. 97: 2870-2874.)
18	7472446	g5983382	4.00E-65	olfactory receptor E3 [Mus musculus] (Krauwurst, D. et al. (1998) Identification of ligands for olfactory receptors by functional expression of a receptor library. Cell 95: 917-926.)

WO 01/66742

PCT/US01/06814

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID MO:	Probability score	GenBank Homolog
19	7472451	G1256393	1.90E-97	taste bud receptor protein TR 641 (Rattus norvegicus) (Thomas, M. B., et al. (1996) Chemoreceptors expressed in taste, olfactory and male reproductive tissues. Gene 178: 1-5.)
20	7472456	G3583374	2.80E-82	olfactory receptor Cf (Mus musculus) (Kraulwurf, D., et al. (1998) Identification of ligands for olfactory receptors by functional expression of a receptor library. Cell 95: 917-926.)
21	7472457	G6178006	3.90E-92	odorant receptor MOR61 (Mus musculus) (Tauboi, A., et al. (1999) Olfactory neurons expressing closely linked and homologous odorant receptor genes tend to project their axons to neighboring glomeruli on the olfactory bulb. J. Neurosci., 19: 8409-8418.)

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Potential Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	536482CD1	99	734		Transmembrane domain: 950-170 Bacterial rhodopsins signature: Y32-R33 Signal peptide: M1-S57 G-protein coupled receptors family 2 signature: E54-G119 TAN3_HUMAN; NETA ISOFORB F0168643: TAN3_S70-L135 (P-value = 4.3e-07) Signal peptide: M1-S63 Melanopsin receptor family signature: L39-P56 Vasopressin V2 receptor signature: P56-S75 Transmembrane domain: L18-P43 Transmembrane receptor (rhodopsin family) domain: G31-Y290 G-protein coupled receptors signature: BL0023191-P120; F181-Y192; L234-L360 Thromboxane receptor signature PR00429: L119-L220 G-protein coupled receptors signature: PR000101[P25968] S1_157; ALJ-R201; P200-R220	HMER PROFLERSCAN SPSCAN PROFLERSCAN BLAST-PRODOM SPSCAN BLIPS-PLIMS BLIPS-PLIMS HMER HMER-PPAM BLIPS-BLOCKS BLIPS-PLIMS BLAST-DOMO
2	1316020CD1	119	S14 T61 S131			
3	2816437CD1	82	Q4 S59 S21 S75 S78	N46		
4	2289894CD1	368	S305	N3 R9		

WO 01/66742

PCT/US01/06814

Table 3 (cont.)

SEQ ID NO:	Invertebrate Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	S376785CD1	153	T20, S46, S110, S116, S144		G-PROTEIN COUPLED RECEPTORS; DM00013 P30731 65-361.I2-S91 PROBABLE G PROTEIN COUPLED RECEPTOR FROM T-CELLS PRECURSOR GUCOCOPPTCCOINDMCCBO GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN SIGNAL ALTERNATIVE SPLICING: PDD61453:C76-S153 Signal cleavage site: K1-C40 Transmembrane domain: T20-L38 7-transmembrane receptor (rhodopsin family): HOMER-PEAM A12-V75 Visual pigments (opsins) retinal binding site: W62-F103 G-PROTEIN COUPLED RECEPTORS: BL00237C:R35-Y41; BL00237D:S67-R83 Rhodopsin-like GPCR super family: PR00237A:K23-S47; PR00237E:K23-L45; PR00237F:K23-L44; PR00237G:Y67-R83 Probable G protein-coupled receptor domain: PR010181:C76-L89	BLAST-DMO BLAST-PRODOM SPSCAN HMMER HOMER-PEAM PROFILESCAN BLIMPS-BLOCKS BLIMPS-PRINTS BLIMPS-PRINTS

WO 01/66742

PCT/US01/06814

Table 3 (cont.)

STO ID NO.	Invertebrate ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	3082743CDL	313	S239 S67 T259 T163 S224 T288	N5 N65	Transmembrane domains: K29-K48; Q100-V118; Y193-A213 7 Transmembrane receptor (rhodopsin family) domain: 641-Y287 G-protein coupled receptors signature BL00237: K90-P129; T279-K295 G-protein coupled receptors signature: E102-L148 Rhodopsin-like GPCR superfamily signature PR00237: L28-K50; R69-K80; L104-I126 K236-R260; K269-K295 Olfactory receptor signature PR00245: K59-K80; Y177-D191; L237-S252 Y271-L282; T288-S302 OLEFACTORY RECEPTOR PROTEIN RECEPTORLIKE SPERMATIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD149621: V246-Q405 G-PROTEIN COUPLED RECEPTORS EMG0013 S28710 15-301; F28-I301 G-protein coupled receptors motif: A110_I126	HMMER HMMER-PPAM BLIMPS-BLOCKS PROTILSCAN BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODOM BLAST-PRODM

WO 01/66742

PCT/US01/06814

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Potential Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	7472361CDL	315	S14 T31 S22 T101 S114 T155 S226 T63 T307 T354 Y351	M12 N61 M89	Transmembrane domains: F78-I96; I182-I210; I239-I262 7 transmembrane receptor (rhodopsin family) Domain: 688-I197; F252-Y338 G-protein coupled receptors signature E400237; E137-E176; P253-Y264; D278-S304 F330-F346 Olfactory receptor signature PR00245; M106-S127; S224-N238; L294-I299 RETINAL GPROTEIN COUPLED RECEPTOR FA1C PD170483; V291-F253 G-PROTEIN COUPLED RECEPTORS DMS0013 645774 18-309; V81-E347 G-protein coupled receptors motif: M157-E173	HMMER HMMER-PFAM BLIMES-BLOCKS BLIMES-PRINTS BLAST-PRODOM BLAST-DMO MOTIFS

WO 01/66742

PCT/US01/06814

Table 3 (cont.)

SEQ ID NO:	Invertebrate Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	747236301	356	T110 S190 S226 S232 S295	N5 N44	Transmembrane Domains: Y37-S55; P60-I84; V204-I223 7 Transmembrane receptor (rhodopsin family) domain: G43-Y294 G-protein coupled receptors signature BL0237: P92-P131; P286-R302 G-protein coupled receptors signature: T164-S154 Olfactory receptor signature F00245: N61-T92; A179-D193; L240-V255 PUTATIVE GPROTEIN COUPLED RECEPTOR EALC FD176483: I247-P109 G-PROTEIN COUPLED RECEPTORS DM00013 S28707 18-306; S23-S302 G-protein coupled receptors motif: L112-L128	IRDBR HMDB-PFAM ELIMPS-BLOCKS PROFILESCAN ELIMPS-PRINTS BLAST-FRDPOM BLAST-DOCK NCI/PS

WO 01/66742

PCT/US01/06814

Table 3 (cont.)

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10	7472364CD	311	T56 S69 T110 N5 R44 S179 T262 T309	N5 R44	Transmembrane domains: F33-I51, I117-I155, I194-I217 7 transmembrane receptor (rhodopsin family) domain: S43-I152; F207-Y293 G-protein coupled receptors signature BL00237: K92-F131; F408-Y219; D233-S459 E285-R301 Olfactory receptor signature PR00246: M61-R82; S179-N193; L238-L254 PUTATIVE GPROTEIN COUPLED RECEPTOR RMAC PDA70483: V246-F308 G-PROTEIN COUPLED RECEPTORS DM00013 645774 18-309: F20-E302 G-protein coupled receptors motif: M113-T128	MEMBER MEMBER-PEAM ELIMPS-BLOCKS ELIMPS-PRINTS BLAST-PRODOM BLAST-DBOM MOTIFS

Table 3 (cont.)

SEQ ID NO	Protein Name	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11	7472434CD1	354	8340 S78 S96 N83 895 E3 T60 864 T116 8316 S316	8340 S78 S96 N83	Transmembrane domain: I218-K235 7 transmembrane receptor (rhodopsin family) domain: N93-Y315 G-protein coupled receptors signature BL02037: K118-P157; T233-Y240; T307-K323 G-protein coupled receptors signature: Y120-L175 Rhodopsin-like GPCR superfamily signature PR00237: T132-I154; L225-E248; K297-K323 Olfactory receptor signature PR00245: R87-L108; Y203-E217; F254-G279 V299-I310; S316-L330 RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L194-K270 G-PROTEIN COUPLED RECEPTORS EM00013 (P23275) 17-306: N92-L330 G-protein coupled receptors motif: T138-I154	HMMER HMMER-PTAM BLIMPS-BLOCKS PROFILESCAN BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODOM BLAST-PCMO MOTIFS

WO 01/66742

PCT/US01/06814

Table 3 (cont.)

SPQ ID No.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13	747243CD1	309	W8 S67 S156 S190 S228 T18 T78 S87 S230 T303	N5 N65	Transmembrane domains: F28-L48; M98-E121 7 transmembrane receptor (rhodopsin family) domain: G41-Y289 G-protein coupled receptors signature E60237: 390-PL29; I281-K297 G-protein coupled receptors signature: Y102-A150 Olfactory receptor signature F801245: K59-K80; Y176-S190; F237-S252 K273-L294; S290-L304 OLFACTORY RECEPTOR PROTEIN RECEPTOR-LIKE PROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L156-L244 G-PROTEIN COUPLED RECEPTORS PM00013 ES1156 18-307: L17-V300	HMMER HMMER-DFAM BLIMPS-BLOCKS PROFILESCAN BLIMPS-PRINTS BLAST-PRODOM BLAST-DMO

WO 01/66742

PCT/US01/06814

Table 3 (cont.)

STG ID No.	Incyte Polyestride ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Potential Signatures, Domains and Motifs	Analytical Methods and Databases
15	7472440CD1	311	S91-S67, T78-S231	N5-N65	Transmembrane domains: R29-R47; Q100-M118; F129-I145 7 Transmembrane receptor (rhodopsin family) domain: G41-V290 G-protein coupled receptors signature: E100237; R90-P129; V282-R288 G-protein coupled receptors signature: F102-G150 Rhodopsin-like GPCR superfamily signature: PR00237; F26-F50; N59-R80; F104-I126; V199-L222; E272-K298 Olfactory receptor signature: PR00245; N59-R80; Y177-S191; F238-G283 RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE PROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD009211; V166-V246 G-PROTEIN COUPLED RECEPTORS: EM00013, S29709 [11-298]; P18-L305 G-protein coupled receptors motif: S310-P126	EMBL HMMER HMMER-PTAM BLASTS-BLOCKS PROFILESCAN BLASTS-PRINTS BLASTS-PRINTS BLAST-PRODOM BLAST-DMG0 KOPRES

WO 01/66742

PCT/US01/06814

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	747243CDL	314	S67 S267 T67 S264 S291	N5 N210	Transmembrane domains: L29-V56; F401-D121; I197-I221 H244-S262 7 transmembrane receptor (rhodopsin family) domain: C41-I290 G-protein coupled receptors signature BL00237: I90-F139; I282-K296 G-protein coupled receptors signature: F102-V150 Olfactory receptor signature PR00745: M59-Q80; P177-D191; P238-G253 V274-L285; S291-R305 RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE PROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY D000921: L166-I245 G-PROTEIN COUPLED RECEPTORS D060013 S51356 18-197: E21-R300 G-protein coupled receptors motif: T110-I126 KTP/GRP binding site (P-loop) S83-T90	HMMER HMMER-PFAM SLIMPS-BLOCKS PROFILESOM SLIMPS-PRINTS BLAST-PRODOM BLAST-DCOMO MOTIFS MOTIFS

WO 01/66742

PCT/US01/06814

Table 3 (cont.)

SEQ ID NO:	Inverte Polyptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Potential Stimulation Domains and Motifs	Analytical Methods and Databases
17	7473445CD1	345	T123 S32 S59 N37 N97 S188 T298 T25 T169 S177 S323 S343		Transmembrane Domains: T51-L83; Y133-D153; M328-L347 7 transmembrane receptor (rhodopsin family) domain: G73-Y322 G-protein coupled receptors signature BL00637: K122-E161; T314-K330 G-protein coupled receptors signature: I134-F178 Rhodopsin-like GPCR superfamily signature W00219: L58-P82; N91-Q112; F136-V158 T231-T254; E304-K330 G-protein coupled receptor signature PRO0245: K91-Q112; Y208-P233; F270-G285; L306-L317; S323-V337 GLFACTORY RECEPTOR PROTEIN RECEPTOR-LIFE MEMBRANE COUPLED TRANSMEMBRANE GPCROTEIN MULTISOME FAMILY PDL49621: S279-K340 G-PROTEIN COUPLED RECEPTORS S000131E03275 17-306: S38-S328 G-PROTEIN COUPLED RECEPTORS MOTIF: T142-V138	EMBL SWISS-PROT PROSITE BLAST-PRODB ELAST-DORO MOTIFS

WO 01/66742

PCT/US01/06814

Table 3 (cont.)

SEQ ID NO.	Inverte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
18	7472445C01	316	S67 S193 T268 T78 S137 T192 S267 S291	R5 N19	Transmembrane domain: L34-A53; E196-T218 / Transmembrane receptor (rhodopsin family) domain: S41-Y280 G-protein coupled receptors signature BRC0237: N90-F129; T382-M298 G-protein coupled receptors signature: F102-T147 Rhodopsin-like GPCR superfamily signature PRC0037: L26-T50; M59-F80; A104-T126 A140-V161; V199-L222; A237-L261 N272-M298 Olfactory receptor signature BRC0245: M59-F80; A177-D191; L238-G253 L274-L285; S291-I308 OLFACTORY RECEPTOR PROTEIN RECEPTOR-LIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD149621: T246-K307 G-PROTEIN COUPLED RECEPTORS DM0001 E12125 17-306; A25-330S	EMBL EMBL-PIR MIMMS-BLOCKS PROFILESCAN MIMMS-PRINTS MIMMS-PRINTS ELAST-FRAGON ELAST-DSMO

WO 01/66742

PCT/US01/06814

Table 3 (cont.)

SEQ ID NO	Incyte Polypeptide Residues	Amino Acid Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
19	7472451CD1 453	S81, T213 T331, T351	N19, N441	Signal peptide: K1-G55 Transmembrane domains: L43-160; I217-7236 7 transmembrane receptor (rhodopsin family) domain: G55-I263 G-protein coupled receptors signature BL0037: R104-P143; T294-K310 G-protein coupled receptors signature: P115-T162 Olfactory receptor signature PR0245: M73-K94; I191-P205; P252-V267 V286-L297; T303-G317 RECEPTOR GUANIDINYL PROTEIN RECEPTOR-LIKE GEM/AFIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY P060921: L150-L259 G-PROTEIN COUPLED RECEPTORS EM00013 S29710 15-301: L43-L315	SESSCAN HMMER HMMER-DFAM BL-MPS-BLOCKS PROFLESSCAN BLMPS-PRIMTS BLAST-PRODOM BLAST-COND

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
20	7472456QD1	323	S49, S67, S188, T230, T40, T291, S320	N5, N65	Transmembrane domains: P25-I45, I57-I85, Y102-D131, I154-Y218 7 transmembrane receptor (rhodopsin family) domain: G41-Y290 G-protein coupled receptors signature: ELMPS-BLOCKS EL00237: R90-P129; T282-Q298 G-protein coupled receptors signature: Y102-A147 Rhodopsin-like GPCR superfamily signature: PRO0237: X59-X80, Y104-Y126; A159-L232, K272-Q298 Olfactory receptor signature: PRO0245: M59-X80, F177-D191; F438-G253, A274-L285; T291-L305 RECEPTOR OLFACTORY PROTEIN RECEPTOR LIKE G-PROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY BDD00971: L156-L245 G-PROTEIN COUPLED RECEPTORS DM00013 P23287 30-309: F17-L305 G-protein coupled receptors motif: T110-L126	HMMER HMMER-PPM ELMPS-BLOCKS PROFILESCAN ELMPS-PRINTS ELMPS-PRINTS ELMPS-PRINTS BLAST-PRODOM ELMPS-DMO MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Potential Signature Sequences, Domains and Motifs	Analytical Methods and Databases
31	7472457CD1	318	S67 T288	N5 N65 N137	<p>Transmembrane domain: V19-V49</p> <p>7 transmembrane receptor (rhodopsin family) domain: S41-V287</p> <p>G-protein coupled receptors signature BL00237: 290-F128; V279-I295</p> <p>G-protein coupled receptors signature: F102-A147</p> <p>Rhodopsin-like GPCR superfamily signature PR00237: V26-T50; M69-R80; F104-I126 V140-R161; M199-L222; A236-R260 K269-I295</p> <p>Olfactory receptor signature PR00245: V271-L282; T288-W302</p> <p>RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE SPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L156-I244</p> <p>G-PROTEIN COUPLED RECEPTORS TM00013 S28710 15-301; L17-W302</p>	<p>HMMER</p> <p>HMMER-PFAM</p> <p>ELIMPS-BLOCKS</p> <p>PROFILESCAN</p> <p>ELIMPS-PRINTS</p> <p>ELIMPS-SHLITS</p> <p>ELAST-PRODOM</p> <p>BLAST-DMMO</p>

WO 01/66742

PCT/US01/06814

Table 4

Polynucleotide SEQ. ID NO.	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
22	536432CBE1	1348	1-174,	687148611 (BRAGN002)	218	950
			812-868	134788276 (PROGNO211)	457	1058
				5364326 (LNDNGT02)	1	436
23	1316020CBE1	1446	1-750,	108761191 (BRATUDR03)	798	1348
			851-810,	558910528 (OPRERG003)	716	1330
			771-832	181645036 (BRAINC003)	1174	1446
				644243291 (BRATN003)	833	1440
				714161611 (BRATN003)	1	445
				817459291 (EPITEX001)	225	613
24	2016437CBE1	1063	155-231,	557537611 (BRGNDT04)	508	769
			1427-1463,	278084726 (OVARNU003)	250	216
			644-800	7017409391	664	1054
				281643791 (BRSTNGT14)	880	1463
				691519591 (FITUDR01)	723	1225
				7017331991	535	722
				294354292 (BRATFUT23)	63	1435
				95743982	1	455
				FL2289894_00001	1	455
				706605041 (BRATN001)	1	512
26	7066050CBE1	2147	629-1512,	FL7066050_00001	11	2132
			1-141	7070662191	1617	2147
27	5376785CBE1	1989	1493-1989,	710153641 (BRATUDR02)	712	1256
			1-63,	7121767391	1242	1989
			875-990,	7082227891	1	602
			157-485,	7081897591	985	1467
			1338-1432,	712170791	485	1122
28	3082743CBE1	942	897-942	GEL_2121229.raw	1	360
				FL3082743_00001	121	942
29	7472361CBE1	948	577-948	GRR_56671985_006	1	948
				GRR_56671985_018	1	948
30	7472363CBE1	1071	554-1071	GRR_56671985_018	1	1071

WO 01/66742

PCT/US01/06814

Table 4 (cont.)

Polynucleotide SEQ ID NO.	Incyte polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
31	74724364CBL	1001	1-55, 435-1001	GNN_#6671982_022	1	1001
32	7472434ACE1	1065	1-356, 1002-1065, 561-594	GNN_#6630753_005	1	1065
33	7472435CE1	963	1-27, 762-963	GNN_#6630753_008	1	963
34	7472438CE1	1101	1-139, 582-712	GNN_#663275_002	1	1101
35	7472439CE1	933		GNN_#663275_006	1	933
36	7472440CE1	936	425-544	GNN_#663275_018	1	936
37	7472443CE1	945	927-945	GNN_#6632708_020	1	945
38	7472445CE1	1041	1008-1041, 61-107	GNN_#6648400_008	1	1041
39	7472446CE1	951	920-951, 578-607	GNN_#6648400_012	1	951
40	7472451CE1	1395	1-87, 1268-1395, 857-1059	GNN_#6648431_006	1	1395
41	7472456CE1	972	862-972	GNN_#6648431_015	1	972
42	7472457CE1	957	916-957	GNN_#6648431_018	1	957

WO 01/66742

PCT/US01/06814

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
22	536487CB1	PROSNUPE1
23	1316020CB1	BRAIN0702
24	28166437CB1	OVARY0203
25	2282834CB1	BRAIN0801
26	7066050CB1	LEUK0102
27	53767850B1	BRAIN0701

WO 01/66742

PCT/US01/06814

Table 6

Library	Vector	Library Description
BRAIN001	ESPORT1	Library was constructed and normalized from 4.88 million independent clones from RNA which was made from brain tissue removed from a 26-year-old Caucasian male during cranioplasty and excision of a cerebral meningeal lesion. Pathology for the associated tumor tissue indicated a grade 4 oligoastrocytoma in the right fronto-parietal part of the brain.
BRAIN002	P1NCY	This large size-fractionated and normalized library was constructed using pooled cDNA generated using mRNA isolated from midbrain, inferior temporal cortex, medulla, and posterior parietal cortex tissue removed from a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Microscopically, the cerebral hemispheres revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. Scattered throughout the cerebral cortex, there were multiple small microscopic areas of cavitation with surrounding gliosis. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly and an enlarged spleen and liver. 2.8 x 10 ⁶ independent clones from this size-selected library were normalized in two rounds using conditions adapted from Sears et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAIN001	P1NCY	Library was constructed using RNA isolated from cerebellar tissue removed from a 70-year-old male. Patient history included chronic obstructive airways disease and left ventricular failure.
LEU00002	P1NCY	Library was constructed using RNA isolated from white blood cells of a 45-year-old female with blood type O+. The donor tested positive for cytomegalovirus (CMV).

WO 01/66742

PCT/US01/06814

Table 6 (cont.)

Library	Vector	Library Description
OVAR0001	pINCY	Library was constructed using RNA isolated from ovarian tumor tissue removed from the left ovary of a 52-year-old mixed ethnicity female during a total abdominal hysterectomy, bilateral salpingo-oophorectomy, peritoneal and lymphatic structure biopsy, regional lymph node excision, and peritoneal tissue destruction. Pathology indicated an invasive grade 3 (of 4) serousneoplastic carcinoma forming a mass in the left ovary. Multiple tumor implants were present on the surface of the left ovary and fallopian tube, right ovary and fallopian tube, posterior surface of the uterus, and cul-de-sac. The endometrium was atrophic. Multiple (2) leiomyomata were identified, one subserosal and 1 intramural. Pathology also indicated a metastatic grade 3 serousneoplastic carcinoma involving the omentum, cul-de-sac peritoneum, left broad ligament peritoneum, and mesentery colon. Patient history included breast cancer, chronic peptic ulcer, and joint pain. Family history included colon cancer, cerebrovascular disease, breast cancer, type II diabetes, esophagus cancer, and depressive disorder.
PRCS0011	pINCY	Library was constructed using DNA isolated from the prostate tissue of a 28-year-old Caucasian male, who died from a self-inflicted gunshot wound.

WO 01/66742

PCT/US01/06814

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.	
ABIPARACEL 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: blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	EST: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
PASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. PASTA comprises at least five functions: fasta, t3sta, fastx, ifastx, and search.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	EST: fasta E value=1.0E-6 Assembled EST: fasta Identity= 95% or greater and Match length=500 bases or greater; fasta E value=1.0E-8 or less Full Length sequences: fasta scores=100 or greater
BLOCKS	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.	Harikeri, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:665-672; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:58-105; and Altschul, T.X. et al. (1997) J. Chem. Inf. Comput. Sci. 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) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 16:3226-3227; Durbin, R. et al. (1998) Cur. World View. in a Nucleic, Cambridge Univ. Press, pp. 1-530.	PFAM hits: Probability value= 1.0E-3 or less Signed peptide hits: Scores= 0 or greater

WO 01/66742

PCT/US01/06814

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gelbukov, M. et al. (1983) CABIOS 1:61-66; Gelbukov, M. et al. (1989) Methods Enzymol. 183:146-159; Beirook, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality scores >CCG-specified "HCGH" 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.	Evling, B. et al. (1998) Genome Res. 8:175-185; Hovig, B. and P. Green (1998) Genome Res. 8:186-194.	
Pump	A Public Revised Assembly Program including SWAT and CrossMatch programs based on efficient implementation of the Smith-Waterman algorithm. Useful in detecting 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 lengths=56 or greater
Conrad	A graphical tool for viewing and editing Emap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presences of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Clewley, J.M. and S. Audile (1997) CABIOS 12:481-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argus (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argus (1996) Protein Sci. 5:563-571.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Southern, 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.	
Modis	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Beirook, 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 01/66742

PCT/US01/06814

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:
 - 5 a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21,
 - b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21,
 - c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and
 - 10 d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-21.
- 15 3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
- 20 5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:22-42.
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.
- 30 9. A method for producing a polypeptide of claim 1, the method comprising:
 - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
 - 35 b) recovering the polypeptide so expressed.

WO 01/66742

PCT/US01/06814

10. An isolated antibody which specifically binds to a polypeptide of claim 1.
11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:
- 5 a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42,
b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42,
c) a polynucleotide sequence complementary to a),
d) a polynucleotide sequence complementary to 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
20 complex is formed between said probe and said target polynucleotide or fragments thereof, and
b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.
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 an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

35

WO 01/66742

PCT/US01/06814

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.
18. A method for treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment the composition of claim 16.
19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:
- 10 a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
b) detecting agonist activity in the sample.
20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.
- 15 21. A method for treating a disease or condition associated with decreased expression of functional GCREC, 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:
- 20 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.
- 25 24. A method for treating a disease or condition associated with overexpression of functional GCREC, comprising administering to a patient in need of such treatment a composition of claim 23.
- 30 25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:
- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
35 b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a

WO 01/66742

PCT/US01/06814

compound that specifically binds to the polypeptide of claim 1.

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

- 5 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
10 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
15 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
20 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
25 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;
c) quantifying the amount of hybridization complex; and
30 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 01/66742

PCT/US01/06814

```

<110> INCYTE GENOMICS, INC.
LAL, Preeti
WANG, Y. Fem
PATTERSON, Chandra
YAO, Monique G.
SHIN, Leo D.
TRIBOLEY, Catherine
LU, Young Aina M.
YUE, Henry
KHAN, Farrah A.
POLICKY, Jennifer L.
SU-YOUNG, Janice
YANG, Junning
HARLAND, Lee
WALSH, Rogerick T.
LO, Terence P.
BOROWSKY, Mark L.

<120> G-PROTEIN COUPLED RECEPTORS
<130> PI-0044 PCT
<140> To Be Assigned
<141> Herewith
<150> 60/186,854; 60/188,384; 60/190,453; 60/190,730
<151> 2000-03-03; 2000-03-10; 2000-03-17; 2000-03-20
<160> 42
<170> PERL Program

<210> 1
<211> 99
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 536482CD1

<400> 1
Met Ala Glu Gly Gly Phe Asp Pro Cys Glu Cys Val Cys Ser His
1 5 10 15
Glu His Ala Met Arg Arg Leu Ile Asn Leu Leu Arg Gln Ser Gln
20 25 30
Ser Tyr Cys Thr Asp Thr Glu Cys Leu Gln Glu Leu Pro Gly Pro
35 40 45
Ser Gly Asp Asn Gly Ile Ser Val Thr Met Ile Leu Val Ala Trp
50 55 60
Met Val Ile Ala Leu Ile Leu Phe Leu Leu Arg Pro Pro Asn Leu
65 70 75
Arg Gly Ser Ser Leu Pro Gly Lys Pro Thr Ser Pro His Asn Gly
80 85 90
Gln Asp Pro Pro Ala Pro Pro Val Asp
95

<210> 2
<211> 139
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1315020CD1

<400> 2
Met Gly His Pro Arg Ala Ile Gln Pro Ser Val Phe Phe Ser Pro

```

WO 01/66742

PCT/US01/06814

```

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

```

```

<210> 3
<211> 82
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 2816437CD1

```

```

<400> 3
Met Gly Lys Thr Pro Ser Glu Ala Gln Asp Ser Leu Val Thr Phe
1           5           10           15
Gln Phe Ala Asp Thr Ser Val Lys Val Ser Trp Glu Thr Ser Ala
20           25           30
Leu Gly Ser Ser Ser Val Val Leu Leu Thr Leu Pro Val Lys Gln
35           40           45
Asn Leu Ser Ser Val Cys Ile Gly Phe His Phe Leu Ser Pro Pro
50           55           60
Glu Glu Trp Lys Ala Thr Ala Gln Ser Leu Leu Met Phe Trp Ser
65           70           75
Glu Arg Ser Leu Lys Val Met
80

```

```

<210> 4
<211> 368
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 2289894CD1

```

```

<400> 4
Met Ala Asn Ser Thr Gly Leu Asn Ala Ser Glu Val Ala Gly Ser
1           5           10           15
Leu Gly Leu Ile Leu Ala Ala Val Val Glu Val Gly Ala Leu Leu
20           25           30
Gly Asn Gly Ala Leu Leu Val Val Val Leu Arg Thr Pro Gly Leu
35           40           45
Arg Asp Ala Leu Tyr Leu Ala His Leu Cys Val Val Asp Leu Leu
50           55           60
Ala Ala Ala Ser Ile Met Pro Leu Gly Leu Leu Ala Ala Pro Pro
65           70           75
Pro Gly Leu Gly Arg Val Arg Leu Gly Pro Ala Pro Cys Arg Ala
80           85           90
Ala Arg Phe Leu Ser Ala Ala Leu Leu Pro Ala Cys Thr Leu Gly
95           100          105

```

WO 01/66742

PCT/US01/06814

Val Ala Ala Leu Gly Leu Ala Arg Tyr Arg Leu Ile Val His Pro
 110 115 120
 Leu Arg Pro Gly Ser Arg Pro Pro Pro Val Leu Val Leu Thr Ala
 125 130 135
 Val Trp Ala Ala Ala Gly Leu Leu Gly Ala Leu Ser Leu Leu Gly
 140 145 150
 Pro Pro Pro Ala Pro Ala Arg Cys Ser Val Leu
 155 160 165
 Ala Gly Gly Leu Gly Pro Phe Arg Pro Leu Trp Ala Leu Leu Ala
 170 175 180
 Phe Ala Leu Pro Ala Leu Leu Leu Gly Ala Tyr Gly Gly Ile
 185 190 195
 Phe Val Val Ala Arg Arg Ala Ala Leu Arg Pro Pro Arg Pro Ala
 200 205 210
 Arg Gly Ser Arg Leu Arg Ser Asp Ser Leu Asp Ser Arg Leu Ser
 215 220 225
 Ile Leu Pro Pro Leu Arg Pro Arg Leu Pro Gly Gly Lys Ala Ala
 230 235 240
 Leu Ala Pro Ala Leu Ala Val Gly Glu Phe Ala Ala Cys Trp Leu
 245 250 255
 Pro Tyr Gly Cys Ala Cys Leu Ala Pro Ala Ala Arg Ala Ala Glu
 260 265 270
 Ala Glu Ala Ala Val Thr Trp Val Ala Tyr Ser Ala Phe Ala Ala
 275 280 285
 His Pro Phe Leu Tyr Gly Leu Leu Glu Arg Pro Val Arg Leu Ala
 290 295 300
 Leu Gly Arg Leu Ser Arg Arg Ala Leu Pro Gly Pro Val Arg Ala
 305 310 315
 Cys Thr Pro Glu Ala Tyr His Pro Arg Ala Leu Leu Glu Cys Leu
 320 325 330
 Glu Arg Pro Pro Glu Gly Pro Ala Val Gly Pro Ser Glu Ala Pro
 335 340 345
 Glu Glu Thr Pro Glu Leu Ala Gly Gly Arg Ser Pro Ala Tyr Glu
 350 355 360
 Gly Pro Pro Glu Ser Ser Leu Ser
 365

<210> 5

<211> 398

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyta ID No. 7066050CD1

<400> 5

Met Glu Ser Gly Leu Leu Arg Pro Ala Pro Val Ser Glu Val Ile
 2 5 10 15
 Val Leu His Tyr Asn Tyr Thr Gly Lys Leu Arg Gly Ala Arg Tyr
 20 25 30
 Glu Pro Gly Ala Gly Leu Arg Ala Asp Ala Val Val Cys Leu Ala
 35 40 45
 Val Cys Ala Phe Ile Val Leu Glu Asn Leu Ala Val Leu Leu Val
 50 55 60
 Leu Gly Arg His Pro Arg Phe His Ala Pro Met Phe Leu Leu Leu
 65 70 75
 Gly Ser Leu Thr Leu Ser Asp Leu Leu Ala Gly Ala Ala Tyr Ala
 80 85 90
 Ala Asn Ile Leu Leu Ser Gly Pro Leu Thr Leu Lys Leu Ser Pro
 95 100 105
 Ala Leu Trp Phe Ala Arg Glu Gly Gly Val Phe Val Ala Leu Thr
 110 115 120
 Ala Ser Val Leu Ser Leu Leu Ala Ile Ala Leu Glu Arg Ser Leu
 125 130 135
 Thr Met Ala Arg Arg Gly Pro Ala Pro Val Ser Ser Arg Gly Arg
 140 145 150

WO 01/66742

PC/US01/06814

```

Thr Leu Ala Met Ala Ala Ala Ala Trp Gly Val Ser Leu Leu Leu
155 160 165
Gly Leu Leu Pro Ala Leu Gly Trp Asn Cys Leu Gly Arg Leu Asp
170 175 180
Ala Cys Ser Thr Val Leu Pro Leu Tyr Ala Lys Ala Tyr Val Leu
185 190 195
Phe Cys Val Leu Ala Phe Val Gly Ile Leu Ala Ala Ile Cys Ala
200 205 210
Leu Tyr Ala Arg Ile Tyr Cys Gln Val Arg Ala Asn Ala Arg Arg
215 220 225
Leu Pro Ala Arg Pro Gly Thr Ala Gly Thr Thr Ser Thr Arg Ala
230 235 240
Arg Arg Lys Pro Arg Ser Leu Ala Leu Leu Arg Thr Leu Ser Val
245 250 255
Val Leu Leu Ala Phe Val Ala Cys Trp Gly Pro Leu Phe Leu Leu
260 265 270
Leu Leu Leu Asp Val Ala Cys Pro Ala Arg Thr Cys Pro Val Leu
275 280 285
Leu Gln Ala Asp Pro Phe Leu Gly Leu Ala Met Ala Asn Ser Leu
290 295 300
Leu Asn Pro Ile Ile Tyr Thr Leu Thr Asn Arg Asp Leu Arg His
305 310 315
Ala Leu Leu Arg Leu Val Cys Cys Gly Arg His Ser Cys Gly Arg
320 325 330
Asp Pro Ser Gly Ser Gln Gln Ser Ala Ser Ala Ala Gln Ala Ser
335 340 345
Gly Gly Leu Arg Arg Cys Leu Pro Pro Gly Leu Asp Gly Ser Phe
350 355 360
Ser Gly Ser Glu Arg Ser Ser Pro Gln Arg Asp Gly Leu Asp Thr
365 370 375
Ser Gly Ser Thr Gly Ser Pro Gly Ala Pro Thr Ala Ala Arg Thr
380 385 390
Leu Val Ser Glu Pro Ala Ala Asp
395

```

```

<210> 6
<211> 153
<212> FRK
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> IncyDe ID No: 5376785CD1

```

```

<400> 6
Met Ile Gly Asp Val Thr Thr Glu Gln Tyr Phe Ala Leu Arg Arg
1 5 10 15
Lys Lys Lys Lys Thr Ile Lys Met Leu Met Leu Val Val Val Leu
20 25 30
Phe Ala Leu Cys Trp Phe Pro Leu Asn Cys Tyr Val Leu Leu Leu
35 40 45
Ser Ser Lys Val Ile Arg Thr Asn Asn Ala Leu Tyr Phe Ala Phe
50 55 60
His Trp Phe Ala Met Ser Ser Thr Cys Tyr Asn Pro Phe Ile Tyr
65 70 75
Cys Trp Leu Asn Glu Asn Phe Arg Ile Glu Leu Lys Ala Leu Leu
80 85 90
Ser Met Cys Gln Arg Pro Pro Lys Pro Gln Glu Asp Arg Gln Pro
95 100 105
Ser Pro Val Pro Ser Phe Arg Val Ala Trp Thr Glu Lys Asn Asp
110 115 120
Gly Gln Arg Ala Pro Leu Ala Asn Asn Leu Leu Pro Thr Ser Gln
125 130 135
Leu Gln Ser Gly Lys Thr Asp Leu Ser Ser Val Glu Pro Ile Val
140 145 150
Thr Met Ser

```

WO 01/66742

PCT/US01/06814

```

<210> 7
<211> 313
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3082743CD1

<400> 7
Met Asp Ser Leu Asn Gln Thr Arg Val Thr Glu Phe Val Phe Leu
1      5      10      15
Gly Leu Thr Asp Asn Arg Val Leu Glu Met Leu Phe Phe Met Ala
20     25     30     35
Phe Ser Ala Ile Ile Tyr Met Leu Thr Leu Ser Gly Asn Ile Leu Ile
40     45     50     55
Ile Ile Ala Thr Val Phe Thr Pro Ser Leu His Thr Pro Met Tyr
60     65     70     75
Phe Phe Leu Ser Asn Leu Ser Phe Ile Asp Ile Cys His Ser Ser
80     85     90     95
Val Thr Val Pro Lys Met Leu Glu Gly Leu Leu Leu Glu Arg Lys
100    105   110   115
Thr Ile Ser Phe Asp Asn Cys Ile Thr Gln Leu Phe Phe Leu His
120    125   130   135
Leu Phe Ala Cys Ala Glu Ile Phe Leu Leu Ile Ile Val Ala Tyr
140    145   150   155
Asp Arg Tyr Val Ala Ile Cys Thr Pro Leu Ile Tyr Pro Asn Val
160    165   170   175
Met Asn Met Arg Val Cys Ile Gln Leu Val Phe Ala Leu Trp Leu
180    185   190   195
Gly Gly Thr Val His Ser Leu Gly Gln Thr Phe Leu Thr Ile Arg
200    205   210   215
Leu Pro Tyr Cys Gly Pro Asn Ile Ile Asp Ser Tyr Phe Cys Asp
220    225   230   235
Val Pro Leu Val Ile Lys Leu Ala Cys Thr Asp Thr Tyr Leu Thr
240    245   250   255
Gly Ile Leu Ile Val Thr Asn Ser Gly Thr Ile Ser Leu Ser Cys
260    265   270   275
Phe Leu Ala Val Val Thr Ser Tyr Met Val Ile Leu Val Ser Leu
280    285   290   295
Arg Lys His Ser Ala Glu Gly Arg Gln Lys Ala Leu Ser Thr Cys
300    305   310   313
Ser Ala His Phe Met Val Val Ala Leu Phe Phe Gly Pro Cys Ile
243   248   253   258
Phe Ile Tyr Thr Arg Pro Asp Thr Ser Phe Ser Ile Asp Lys Val
260   265   270   275
Val Ser Val Phe Tyr Thr Val Val Thr Pro Leu Leu Asn Pro Phe
280   285   290   295
Ile Tyr Thr Leu Arg Asn Glu Glu Val Lys Ser Ala Met Lys Gln
295   300   305   310
Leu Arg Gln Arg Gln Val Phe Phe Thr Lys Ser Tyr Thr
305   310   313

```

```

<210> 8
<211> 315
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472361CD1

```

```

<400> 8
Met Gly Asp Trp Asn Asn Ser Asp Ala Val Glu Pro Ile Phe Ile
1      5      10      15
Leu Arg Gly Phe Pro Gly Leu Glu Tyr Val His Ser Trp Leu Ser
20     25     30

```

WO 01/66742

PCT/US01/06814

```

Ile Leu Phe Cys Leu Ala Tyr Leu Val Ala Phe Met Gly Asn Val
35 40 45
Thr Ile Leu Ser Val Ile Trp Ile Glu Ser Ser Leu His Gln Pro
50 55 60
Met Tyr Tyr Phe Ile Ser Ile Leu Ala Val Asn Asp Leu Gly Met
65 70 75
Ser Leu Ser Thr Leu Pro Thr Met Leu Ala Val Leu Trp Leu Asp
80 85 90
Ala Pro Glu Ile Gln Ala Ser Ala Cys Tyr Ala Gln Leu Phe Phe
95 100 105
Ile His Thr Phe Thr Phe Leu Glu Ser Ser Val Leu Leu Ala Met
110 115 120
Ala Phe Asp Arg Phe Val Ala Ile Cys His Pro Leu His Tyr Pro
125 130 135
Thr Ile Leu Thr Asn Ser Val Ile Gly Lys Ile Gly Leu Ala Cys
140 145 150
Leu Leu Arg Ser Leu Gly Val Val Leu Pro Thr Pro Leu Leu Leu
155 160 165
Arg His Tyr His Tyr Cys His Gly Asn Ala Leu Ser His Ala Phe
170 175 180
Cys Leu His Gln Asp Val Leu Arg Leu Ser Cys Thr Asp Ala Arg
185 190 195
Thr Asn Ser Ile Tyr Gly Leu Cys Val Val Ile Ala Thr Leu Gly
200 205 210
Val Asp Ser Ile Phe Ile Leu Leu Ser Tyr Val Leu Ile Leu Asn
215 220 225
Thr Val Leu Asp Ile Ala Ser Arg Glu Glu Gln Leu Lys Ala Leu
230 235 240
Asn Thr Cys Val Ser His Ile Cys Val Val Leu Ile Phe Phe Val
245 250 255
Pro Val Ile Gly Val Ser Met Val His Arg Phe Gly Lys His Leu
260 265 270
Ser Pro Ile Val His Ile Leu Met Ala Asp Ile Tyr Leu Leu Leu
275 280 285
Pro Pro Val Leu Asn Pro Ile Val Tyr Ser Val Arg Thr Lys Gln
290 295 300
Ile Arg Leu Gly Ile Leu His Lys Phe Val Leu Arg Arg Arg Phe
305 310 315

```

<210> 9

<211> 356

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472363CD1

<400> 9

```

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

```

WO 01/66742

PCT/US01/06814

125 130 135
 Ala His Glu Ile Asn Tyr Gly Ala Cys Val Ala Gln Met Phe Leu
 149 145 150
 Ile His Ala Phe Thr Gly Met Glu Ala Glu Val Leu Leu Ala Met
 155 160 165
 Ala Phe Asp Arg Tyr Val Ala Ile Cys Ala Pro Leu His Tyr Ala
 170 175 180
 Thr Ile Leu Thr Ser Leu Val Leu Val Gly Ile Ser Met Cys Ile
 185 190 195
 Val Ile Arg Pro Val Leu Leu Thr Leu Pro Met Val Tyr Leu Ile
 200 205 210
 Tyr Arg Leu Pro Phe Cys Gln Ala His Ile Ile Ala His Ser Tyr
 215 220 225
 Cys Glu His Met Gly Ile Ala Lys Leu Ser Cys Gly Asn Ile Arg
 230 235 240
 Ile Asn Gly Ile Tyr Gly Leu Phe Val Val Ser Phe Phe Val Leu
 245 250 255
 Asn Leu Val Leu Ile Gly Ile Ser Tyr Val Tyr Ile Leu Arg Ala
 260 265 270
 Val Phe Arg Leu Pro Ser His Asp Ala Gln Leu Lys Ala Leu Ser
 275 280 285
 Thr Cys Gly Ala His Val Gly Val Ile Cys Val Phe Tyr Ile Pro
 290 295 300
 Ser Val Phe Ser Phe Leu Thr His Arg Phe Gly His Gln Ile Pro
 305 310 315
 Gly Tyr Ile His Ile Leu Val Ala Asn Leu Tyr Leu Ile Ile Pro
 320 325 330
 Pro Ser Leu Asn Pro Ile Ile Tyr Gly Val Arg Thr Lys Gln Ile
 335 340 345
 Arg Glu Arg Val Leu Tyr Val Phe Thr Lys Lys
 350 355

<210> 10

<211> 311

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<222> Incyte ID No: 7472364CD1

<400> 10

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

WO 01/66742

PC/T/US01/06814

```

185          190          195
Ile Asn Gly Ile Tyr Gly Leu Phe Val Val Ser Phe Phe Val Leu
200          205          210
Asn Leu Val Leu Ile Gly Ile Ser Tyr Val Tyr Ile Leu Arg Ala
215          220          225
Val Phe Arg Leu Pro Ser His Asp Ala Gln Leu Lys Ala Leu Ser
230          235          240
Thr Cys Gly Ala His Val Gly Val Ile Cys Val Phe Tyr Ile Pro
245          250          255
Ser Val Phe Ser Phe Leu Thr His Arg Phe Gly His Gln Ile Pro
260          265          270
Gly Tyr Ile His Ile Leu Val Ala Asn Leu Tyr Leu Ile Ile Pro
275          280          285
Pro Ser Leu Asn Pro Ile Ile Tyr Gly Val Arg Thr Lys Gln Ile
290          295          300
Arg Glu Arg Val Leu Tyr Val Phe Thr Lys Lys
305          310

```

```

<210> 11
<211> 354
<212> PRF
<213> Homo sapiens

```

```

<220>
<221> misc feature
<223> Incyte 3D No: 7472434CD1

```

```

<400> 11
Met Arg Ser Leu Lys Ala Gly Gly Lys Gln Thr Val Tyr Val Ala
1          5          10          15
Gly Glu Gln Glu Ala Gly Ile Pro Asp Ala Gly Leu Ser Arg Gly
20          25          30
Glu Val Arg Ala Ala Leu His Gly Asp Gly Gly His Leu Gly Gln
35          40          45
Thr Thr Ala Ser Pro Thr Ala Pro Phe Ala Lys Leu Val Thr Thr
50          55          60
Asp Arg Thr Ser Thr Arg Phe Val Pro Gly Phe Pro Pro Arg Val
65          70          75
Thr Ser Leu Ser Val Ser Phe Leu Leu Gln Ser Asn Met Glu Ala
80          85          90
Arg Asn Asn Leu Ser Leu Met Asp Ile Cys Gly Thr Ser Ser Phe
95          100          105
Val Pro Leu Met Leu Asp Asn Phe Leu Glu Thr Gln Arg Thr Ile
110          115          120
Ser Phe Pro Gly Cys Ala Leu Gln Met Tyr Leu Thr Leu Ala Leu
125          130          135
Gly Ser Thr Glu Cys Leu Leu Leu Ala Val Met Ala Tyr Asp Arg
140          145          150
Tyr Val Ala Ile Cys Gln Pro Leu Arg Tyr Pro Glu Leu Met Ser
155          160          165
Gly Gln Thr Cys Met Gln Met Ala Ala Leu Ser Trp Gly Thr Gly
170          175          180
Phe Ala Asn Ser Leu Leu Gln Ser Ile Leu Val Trp His Leu Pro
185          190          195
Phe Cys Gly His Val Ile Asn Tyr Phe Tyr Glu Ile Leu Ala Val
200          205          210
Leu Lys Leu Ala Cys Gly Asp Ile Ser Leu Asn Ala Leu Ala Leu
215          220          225
Met Val Ala Thr Ala Val Leu Thr Leu Ala Pro Leu Leu Leu Ile
230          235          240
Cys Leu Ser Tyr Leu Phe Ile Leu Ser Ala Ile Leu Arg Val Pro
245          250          255
Ser Ala Ala Gly Arg Cys Lys Ala Phe Ser Thr Cys Ser Ala His
260          265          270
Arg Thr Val Val Val Val Phe Tyr Gly Thr Ile Ser Phe Met Tyr
275          280          285
Phe Lys Pro Lys Ala Lys Asp Pro Asn Val Asp Lys Thr Val Ala

```

WO 01/66742 PCT/US01/06814

290 295 300
 Leu Phe Tyr Gly Val Val Thr Pro Ser Leu Asn Pro Ile Ile Tyr
 305 310 315
 Ser Leu Arg Asn Ala Glu Val Lys Ala Ala Val Leu Thr Leu Leu
 320 325 330
 Arg Gly Gly Leu Leu Ser Arg Lys Ala Ser His Cys Tyr Cys Cys
 335 340 345
 Pro Leu Pro Leu Ser Ala Gly Ile Gly 350

<210> 12
 <211> 319
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472435CD1

<400> 12
 Met Glu Lys Ala Asn Glu Thr Ser Pro Val Met Gly Phe Val Leu
 1 5 10 15
 Leu Arg Leu Ser Ala His Pro Glu Leu Glu Lys Thr Phe Phe Val
 20 25 30
 Leu Ile Leu Leu Met Tyr Leu Val Ile Leu Leu Gly Asn Gly Val
 35 40 45
 Leu Ile Leu Val Thr Ile Leu Asp Ser Arg Leu His Thr Pro Met
 50 55 60
 Tyr Phe Phe Leu Gly Asn Leu Ser Phe Leu Asp Ile Cys Phe Thr
 65 70 75
 Thr Ser Ser Val Pro Leu Val Leu Asp Ser Phe Leu Thr Pro Gln
 80 85 90
 Glu Thr Ile Ser Phe Ser Ala Cys Ala Val Gln Met Ala Leu Ser
 95 100 105
 Phe Ala Met Ala Gly Thr Glu Cys Leu Leu Leu Ser Met Met Ala
 110 115 120
 Phe Asp Arg Tyr Val Ala Ile Cys Asn Pro Leu Arg Tyr Ser Val
 125 130 135
 Ile Met Ser Lys Ala Ala Tyr Met Pro Met Ala Ala Ser Ser Trp
 140 145 150
 Ala Ile Ciy Gly Ala Ala Ser Val Val His Thr Ser Leu Ala Ile
 155 160 165
 Ctr Leu Pro Phe Cys Gly Asp Asn Val Ile Asn His Phe Thr Cys
 170 175 180
 Glu Ile Leu Ala Val Leu Lys Leu Ala Cys Ala Asp Ile Ser Ile
 185 190 195
 Asn Val Ile Ser Met Glu Val Thr Asn Val Ile Phe Leu Gly Val
 200 205 210
 Pro Val Leu Phe Ile Ser Phe Ser Tyr Val Phe Ile Ile Thr Thr
 215 220 225
 Ile Leu Arg Ile Pro Ser Ala Glu Gly Arg Lys Lys Val Phe Ser
 230 235 240
 Thr Cys Ser Ala His Leu Thr Val Val Ile Val Phe Tyr Gly Thr
 245 250 255
 Leu Phe Phe Met Tyr Gly Lys Pro Lys Ser Lys Asp Ser Met Gly
 260 265 270
 Ala Asp Lys Glu Asp Leu Ser Asp Lys Leu Ile Pro Leu Phe Tyr
 275 280 285
 Gly Val Val Thr Pro Met Leu Asn Pro Ile Ile Tyr Ser Leu Arg
 290 295 300
 Asn Lys Asp Val Lys Ala Ala Val Arg Arg Leu Leu Arg Pro Lys
 305 310 315
 Gly Phe Thr Gln

<210> 13
 <211> 309

WO 01/66742

PCT/US01/06814

<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472438CD1

<220>
<221> unsure
<222> 125-126
<223> unknown or other

```

<400> 13
Met Ala Ala Gly Asn His Ser Thr Val Thr Glu Phe Ile Leu Lys
1      5      10      15
Gly Leu Thr Lys Arg Ala Asp Leu Gln Leu Pro Leu Phe Leu Leu
20     25     30     35     40     45
Phe Leu Gly Ile Cys Leu Asn Ser Gln Leu His Thr Pro Met Tyr
50     55     60     65     70
Val Ile Thr Pro Lys Met Leu Val Asn Phe Val Ser Glu Lys Asn
80     85     90     95
Ile Ile Ser Tyr Ala Gly Cys Met Ser Gln Leu Tyr Phe Phe Leu
100    105    110    115
Val Phe Val Ile Ala Glu Cys Tyr Met Leu Thr Val Met Ala Tyr
120    125    130    135
Asp Arg Tyr Val Xaa Xaa Cys His Pro Leu Leu Tyr Asn Ile Ile
140    145    150    155
Met Ser His His Phe Cys Leu Leu Leu Val Ala Val Val Tyr Ala
160    165    170    175
Ile Gly Leu Ile Gly Ser Thr Ile Glu Thr Gly Leu Met Leu Lys
180    185    190    195
Leu Pro Tyr Cys Glu His Leu Ile Ser His Tyr Phe Cys Asp Ile
200    205    210    215
Leu Pro Leu Met Lys Leu Ser Cys Ser Ser Thr Tyr Asp Val Glu
220    225    230    235
Met Thr Val Phe Phe Ser Ala Gly Phe Asn Ile Ile Val Thr Ser
240    245    250    255
Leu Thr Val Leu Val Ser Tyr Thr Phe Ile Leu Ser Ser Ile Leu
260    265    270    275
Gly Ile Ser Thr Thr Glu Gly Arg Ser Lys Ala Phe Ser Thr Cys
280    285    290    295
Ser Ser His Leu Ala Ala Val Gly Met Phe Tyr Gly Ser Thr Ala
300    305
Phe Met Tyr Leu Lys Pro Ser Thr Ile Ser Ser Leu Thr Gln Glu
310    315    320
Asx Val Ala Ser Val Phe Tyr Thr Thr Val Ile Pro Met Leu Asn
325    330    335
Pro Leu Ile Tyr Ser Leu Arg Asn Lys Glu Val Lys Ala Ala Val
340    345    350
Gln Lys Thr Leu Arg Gly Lys Leu Phe
355

```

<210> 14
<211> 310
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472439CD1

<400> 14
Met Ala Ala Lys Asn Ser Ser Val Thr Glu Phe Ile Leu Glu Gly

WO 01/66742

PCT/US01/06814

```

1           5           10           15
Leu Thr His Gln Pro Gly Leu Arg Ile Pro Leu Phe Phe Leu Phe
20
Leu Gly Phe Tyr Thr Val Thr Val Val Gly Asn Leu Gly Leu Ile
35
Thr Leu Ile Gly Leu Asn Ser His Leu His Thr Pro Met Tyr Phe
50
Phe Leu Phe Asn Leu Ser Leu Ile Asp Phe Cys Phe Ser Thr Thr
65
Ile Thr Pro Lys Met Leu Met Ser Phe Val Ser Arg Lys Asn Ile
80
Ile Ser Phe Thr Gly Cys Met Thr Gln Leu Phe Phe Phe Cys Phe
95
Phe Val Val Ser Glu Ser Phe Ile Leu Ser Ala Met Ala Tyr Asp
110
Arg Tyr Val Ala Ile Cys Asn Pro Leu Leu Tyr Thr Val Thr Met
125
Ser Cys Gln Val Cys Leu Leu Leu Leu Gly Ala Tyr Gly Met
140
Gly Phe Ala Gly Ala Met Ala His Thr Gly Ser Ile Met Asn Leu
155
Thr Phe Cys Ala Asp Asn Leu Val Asn His Phe Met Cys Asp Ile
170
Leu Pro Leu Leu Glu Leu Ser Cys Asn Ser Ser Tyr Met Asn Glu
185
Leu Val Val Phe Ile Val Val Ala Val Asp Val Gly Met Pro Ile
200
Val Thr Val Phe Ile Ser Tyr Ala Leu Ile Leu Ser Ser Ile Leu
215
His Asn Ser Ser Thr Glu Gly Arg Ser Lys Ala Phe Ser Thr Cys
230
Ser Ser His Ile Ile Val Val Ser Leu Phe Phe Gly Ser Gly Ala
245
Phe Met Tyr Leu Lys Pro Leu Ser Ile Leu Pro Leu Glu Gln Gly
260
Lys Val Ser Ser Leu Phe Tyr Thr Ile Ile Val Pro Val Leu Asn
275
Pro Leu Ile Tyr Ser Leu Arg Asn Lys Asp Val Lys Val Ala Leu
290
Arg Arg Thr Leu Gly Arg Lys Ile Phe Ser
305

```

<210> 15

<211> 311

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472440CD1

<400> 15

```

Met Ala Ala Glu Asn Ser Ser Phe Val Thr Gln Phe Ile Leu Ala
1           5           10           15
Gly Leu Thr Asp Gln Pro Gly Val Gln Ile Pro Leu Phe Phe Leu
20
Phe Leu Gly Phe Tyr Val Val Thr Val Val Gly Asn Leu Gly Leu
35
Ile Thr Leu Ile Arg Leu Asn Ser His Leu His Thr Pro Met Tyr
50
Phe Phe Leu Tyr Asn Leu Ser Phe Ile Asp Phe Cys Tyr Ser Ser
65
Val Ile Thr Pro Lys Met Leu Met Ser Phe Val Leu Lys Lys Asn
80
Ser Ile Ser Tyr Ala Gly Cys Met Thr Gln Leu Phe Phe Phe Leu
95
Phe Phe Val Val Ser Glu Ser Phe Ile Leu Ser Ala Met Ala Tyr

```

WO 01/66742

PCT/US01/06814

110 115 120
 Asp Arg Tyr Val Ala Ile Cys Asn Pro Leu Leu Tyr Met Val Thr
 125 130 135
 Met Ser Pro Gln Val Cys Phe Leu Leu Leu Leu Gly Val Tyr Gly
 140 145 150
 Met Gly Phe Ala Gly Ala Met Ala His Thr Ala Cys Met Met Gly
 155 160 165
 Val Thr Phe Cys Ala Asn Asn Leu Val Asn His Tyr Met Cys Asp
 170 175 180
 Ile Leu Pro Leu Leu Glu Cys Ala Cys Thr Ser Thr Tyr Val Asn
 185 190 195
 Glu Leu Val Val Phe Val Val Val Gly Ile Asp Ile Gly Val Pro
 200 205 210
 Thr Val Thr Ile Phe Ile Ser Tyr Ala Leu Ile Leu Ser Ser Ile
 215 220 225
 Phe His Ile Asp Ser Thr Glu Gly Arg Ser Lys Ala Phe Ser Thr
 230 235 240
 Cys Ser Ser His Ile Ile Ala Val Ser Leu Phe Phe Gly Ser Gly
 245 250 255
 Ala Phe Met Tyr Leu Lys Pro Phe Ser Leu Leu Ala Met Asn Gln
 260 265 270
 Gly Lys Val Ser Ser Leu Phe Tyr Thr Thr Val Val Pro Met Leu
 275 280 285
 Asn Pro Leu Ile Tyr Ser Leu Arg Asn Lys Asp Val Lys Val Ala
 290 295 300
 Leu Lys Lys Ile Leu Asn Lys Asn Ala Phe Ser
 305 310

<210> 16
 <211> 314
 <212> PRF
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472443CD1

<400> 16
 Met Ala Lys Asn Asn Leu Thr Arg Val Thr Glu Phe Ile Leu Met
 1 5 10 15
 Gly Phe Met Asp His Pro Lys Leu Glu Ile Pro Leu Phe Leu Val
 20 25 30
 Phe Leu Ser Phe Tyr Leu Val Thr Leu Leu Gly Asn Val Gly Met
 35 40 45
 Ile Met Leu Ile Gln Val Asp Val Lys Leu Tyr Thr Pro Met Tyr
 50 55 60
 Phe Phe Leu Ser His Leu Ser Leu Leu Asp Ala Cys Tyr Thr Ser
 65 70 75
 Val Ile Thr Pro Gln Ile Leu Ala Thr Leu Ala Thr Gly Lys Thr
 80 85 90
 Val Ile Ser Tyr Gly His Cys Ala Ala Gln Phe Phe Leu Phe Thr
 95 100 105
 Ile Cys Ala Gly Thr Glu Cys Phe Leu Leu Ala Val Met Ala Tyr
 110 115 120
 Asp Arg Tyr Ala Ala Ile Arg Asn Pro Leu Leu Tyr Thr Val Ala
 125 130 135
 Met Asn Pro Arg Leu Cys Trp Ser Leu Val Val Gly Ala Tyr Val
 140 145 150
 Cys Gly Val Ser Gly Ala Ile Leu Arg Thr Thr Cys Thr Phe Thr
 155 160 165
 Leu Ser Phe Cys Lys Asp Asn Gln Ile Asn Phe Phe Phe Cys Asp
 170 175 180
 Leu Pro Pro Leu Leu Lys Leu Ala Cys Ser Asp Thr Ala Asn Ile
 185 190 195
 Glu Ile Val Ile Ile Phe Phe Gly Asn Phe Val Ile Leu Ala Asn
 200 205 210
 Ala Ser Val Ile Leu Ile Ser Tyr Leu Leu Ile Ile Lys Thr Ile

WO 01/66742

PCT/US01/06814

215	220	235
Leu Lys Val Lys Ser Ser Gly Gly Arg Ala Lys Thr Phe Ser Thr		
230	235	240
Cys Ala Ser His Ile Thr Ala Val Ala Leu Phe Phe Gly Ala Leu		
245	250	255
Ile Phe Met Tyr Leu Gln Ser Gly Ser Gly Lys Ser Leu Glu Gln		
260	265	270
Asp Lys Val Val Ser Val Phe Tyr Thr Val Val Ile Pro Met Leu		
275	280	285
Asn Pro Leu Ile Tyr Ser Leu Arg Asn Lys Asp Val Lys Asp Ala		
290	295	300
Phe Arg Lys Val Ala Arg Arg Leu Glu Val Ser Leu Ser Met		
305	310	

<210> 17
 <211> 346
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> incyte ID No: 7472445CD1

<400> 17
 Met Asn Ala Ala Gln Ser His Tyr Pro Lys Arg Thr Asn Ala Gly
 1 5 10 15
 Thr Gly Asn Gln Ile Ser His Val Leu Thr Cys Lys Gln Ala Lys
 20 25 30
 Ile Ser Met Gly Glu Glu Asn Gln Thr Phe Val Ser Lys Phe Ile
 35 40 45
 Phe Leu Gly Leu Ser Gln Asp Leu Gln Thr Gln Ile Leu Leu Phe
 50 55 60
 Ile Leu Phe Leu Ile Ile Tyr Leu Leu Thr Val Leu Gly Asn Gln
 65 70 75
 Leu Ile Ile Ile Leu Ile Phe Leu Asp Ser Arg Leu His Thr Pro
 80 85 90
 Met Tyr Phe Phe Leu Arg Asn Leu Ser Phe Ala Asp Leu Cys Phe
 95 100 105
 Ser Thr Ser Ile Val Pro Gln Val Leu Val His Phe Leu Val Lys
 110 115 120
 Arg Lys Thr Ile Ser Phe Tyr Gly Cys Met Thr Gln Ile Ile Val
 125 130 135
 Phe Leu Leu Val Gly Cys Thr Glu Cys Ala Leu Leu Ala Val Met
 140 145 150
 Ser Tyr Asp Arg Tyr Val Ala Val Cys Lys Pro Leu Tyr Tyr Ser
 155 160 165
 Thr Ile Met Thr Gln Arg Val Cys Leu Trp Leu Ser Phe Arg Ser
 170 175 180
 Trp Ala Ser Gly Ala Leu Val Ser Leu Val Asp Thr Ser Phe Thr
 185 190 195
 Phe His Leu Pro Tyr Trp Gly Gln Asn Ile Ile Asn His Tyr Phe
 200 205 210
 Cys Glu Pro Pro Ala Leu Leu Lys Leu Ala Ser Ile Asp Thr Tyr
 215 220 225
 Ser Thr Glu Met Ala Ile Phe Ser Met Gly Val Val Ile Leu Leu
 230 235 240
 Ala Pro Val Ser Leu Ile Leu Gly Ser Tyr Trp Asn Ile Ile Ser
 245 250 255
 Thr Val Ile Gln Met Gln Ser Gly Glu Gly Arg Leu Lys Ala Phe
 260 265 270
 Ser Thr Cys Gly Ser His Leu Ile Val Val Val Leu Phe Tyr Gly
 275 280 285
 Ser Gly Ile Phe Thr Tyr Met Arg Pro Asn Ser Lys Thr Thr Lys
 290 295 300
 Glu Leu Asp Lys Met Ile Ser Val Phe Tyr Thr Ala Val Thr Pro
 305 310 315
 Met Leu Asn Pro Ile Ile Tyr Ser Leu Arg Asn Lys Asp Val Lys

WO 01/66742 PCT/US01/06814

320 325 330
 Gly Ala Leu Arg Lys Leu Val Gly Arg Lys Cys Phe Ser His Arg
 335 340 345
 Gln

<210> 18
 <211> 316
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472446CD1

<400> 18
 Met Glu Leu Trp Asn Phe Thr Leu Gly Ser Gly Phe Ile Leu Val
 1 5 10 15
 Gly Ile Leu Asn Asp Ser Gly Ser Pro Glu Leu Leu Cys Ala Thr
 20 25 30
 Ile Thr Ile Leu Tyr Leu Leu Ala Leu Ile Ser Asn Gly Leu Leu
 35 40 45
 Leu Leu Ala Ile Thr Met Glu Ala Arg Leu His Met Pro Met Tyr
 50 55 60
 Leu Leu Leu Gly Gln Leu Ser Leu Met Asp Leu Leu Phe Thr Ser
 65 70 75
 Val Val Thr Pro Lys Ala Leu Ala Asp Phe Leu Arg Arg Glu Asn
 80 85 90
 Thr Ile Ser Phe Gly Gly Cys Ala Leu Gln Met Phe Leu Ala Leu
 95 100 105
 Thr Met Gly Gly Ala Glu Asp Leu Leu Leu Ala Phe Met Ala Tyr
 110 115 120
 Asp Arg Tyr Val Ala Ile Cys His Pro Leu Thr Tyr Met Thr Leu
 125 130 135
 Met Ser Ser Arg Ala Cys Trp Leu Met Val Ala Thr Ser Trp Ile
 140 145 150
 Leu Ala Ser Leu Ser Ala Leu Ile Tyr Thr Val Tyr Thr Met His
 155 160 165
 Tyr Pro Phe Cys Arg Ala Gln Glu Ile Arg His Leu Leu Cys Glu
 170 175 180
 Ile Pro His Leu Leu Lys Val Ala Cys Ala Asp Thr Ser Arg Tyr
 185 190 195
 Glu Leu Met Val Tyr Val Met Gly Val Thr Phe Leu Ile Pro Ser
 200 205 210
 Leu Ala Ala Ile Leu Ala Ser Tyr Thr Gln Ile Leu Leu Thr Val
 215 220 225
 Leu His Met Pro Ser Asn Glu Gly Arg Lys Lys Ala Leu Val Thr
 230 235 240
 Cys Ser Ser His Leu Thr Val Val Gly Met Phe Tyr Gly Ala Ala
 245 250 255
 Thr Phe Met Tyr Val Leu Pro Ser Ser Phe His Ser Thr Arg Gln
 260 265 270
 Asp Asn Ile Ile Ser Val Phe Tyr Thr Ile Val Thr Pro Ala Leu
 275 280 285
 Asn Pro Leu Ile Tyr Ser Leu Arg Asn Lys Glu Val Met Arg Ala
 290 295 300
 Leu Arg Arg Val Leu Gly Lys Tyr Met Leu Pro Ala His Ser Thr
 305 310 315
 Leu

<210> 13
 <211> 453
 <212> PRT
 <213> Homo sapiens

<220>

WO 01/66742

PCT/US01/06814

<221> misc feature
<225> Incyte ID No: 7472451C01

```

<400> 19
Met Leu Tyr Lys Tyr Leu Glu Arg Asp Val Asn Ser Lys Glu Leu
1 5 10 15
Gln Ser Gly Asn Gln Thr Ser Val Ser His Phe Ile Leu Val Gly
20 25 30
Leu His His Pro Pro Gln Leu Gly Ala Pro Leu Phe Leu Ala Phe
35 40 45
Leu Val Ile Tyr Leu Leu Thr Val Ser Gly Asn Gly Leu Ile Ile
50 55 60
Leu Thr Val Leu Val Asp Ile Arg Leu His Arg Pro Met Cys Leu
65 70 75
Phe Leu Cys His Leu Ser Phe Leu Asp Met Thr Ile Ser Cys Ala
80 85 90
Ile Val Pro Lys Met Leu Ala Gly Phe Leu Leu Gly Ser Arg Ile
95 100 105
Ile Ser Phe Gly Gly Cys Val Ile Gln Leu Phe Ser Phe His Phe
110 115 120
Leu Gly Cys Thr Glu Cys Phe Leu Tyr Thr Leu Met Ala Tyr Asp
125 130 135
Arg Phe Leu Ala Ile Cys Lys Pro Leu His Tyr Ala Thr Ile Met
140 145 150
Thr His Arg Val Cys Asn Ser Leu Ala Leu Gly Thr Trp Leu Gly
155 160 165
Gly Thr Ile His Ser Leu Phe Gln Thr Ser Phe Val Phe Arg Leu
170 175 180
Pro Phe Cys Gly Pro Asn Arg Val Asp Tyr Ile Phe Cys Asp Ile
185 190 195
Pro Ala Met Leu Arg Leu Ala Cys Ala Asp Thr Ala Ile Asn Glu
200 205 210
Leu Val Thr Phe Ala Asp Ile Gly Phe Leu Ala Leu Thr Cys Phe
215 220 225
Met Leu Ile Leu Thr Ser Tyr Gly Tyr Ile Val Ala Ala Ile Leu
230 235 240
Arg Ile Pro Ser Ala Asp Gly Arg Arg Asn Ala Phe Ser Thr Cys
245 250 255
Ala Ala His Leu Thr Val Val Ile Val Tyr Tyr Val Pro Cys Thr
260 265 270
Phe Ile Tyr Leu Arg Pro Cys Ser Gln Gln Pro Leu Asp Gly Val
275 280 285
Val Ala Val Phe Tyr Thr Val Ile Thr Pro Leu Leu Asn Ser Ile
290 295 300
Ile Tyr Thr Leu Cys Asn Lys Glu Met Lys Ala Ala Leu Gln Arg
305 310 315
Leu Gly Gly His Lys Glu Glu Val Glu Glu Ile Glu Leu Gly His
320 325 330
Thr Thr Val Gln Gly His Ser Leu Ala Thr Gln Gly Gln Gln Gly
335 340 345
Pro Arg His Phe Gly His His Asp Ser Glu Glu Pro Gln Val Asn
350 355 360
Glu Gly Gln Ile Gly Glu Glu Val Val Leu Gly Gly Val Glu Val
365 370 375
Arg Val His Pro Asp His Gln Gln Asp Glu Glu Val Pro Gln His
380 385 390
Asn Leu Lys Lys Asn Tyr Met Thr Ile Ser Ala Asp Gly Glu Lys
395 400 405
Ala Leu Asp Asn Ile Arg His Pro Leu Ile Lys Thr Leu Asn Asn
410 415 420
Leu Glu Val Lys Gly Asp Phe Leu Asn Leu Met Lys Asp Val Tyr
425 430 435
Glu Asn Pro Thr Pro Asn Leu Ser Lys Tyr Pro Glu Lys Leu Asn
440 445 450
Ala Phe Pro

```

WO 01/66742

PC/US01/06814

<210> 20
<211> 323
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472456CD1

```

<400> 20
Met Asn Pro Glu Asn Trp Thr Gln Val Thr Ser Phe Val Leu Leu
1      5      10      15
Cly Phe Pro Ser Ser His Leu Ile Gln Phe Leu Val Phe Leu Gly
26      30      35      40      45
Leu Met Val Thr Tyr Ile Val Thr Ala Thr Gly Lys Leu Leu Ile
35      40      45      50      55      60
Ile Val Leu Ser Trp Ile Asp Gln Arg Leu His Ile Gln Met Tyr
50      55      60      65      70      75
Phe Phe Leu Arg Asn Phe Ser Phe Leu Glu Leu Leu Leu Val Thr
65      70      75      80      85      90
Val Val Val Pro Lys Met Leu Val Val Ile Leu Thr Gly Asp His
95      100      105      110      115      120
Thr Ile Ser Phe Val Ser Cys Ile Ile Gln Ser Tyr Leu Tyr Phe
105      110      115      120      125      130
Phe Leu Gly Thr Thr Asp Phe Phe Leu Leu Ala Val Met Ser Leu
110      115      120      125      130      135
Asp Arg Tyr Leu Ala Ile Cys Arg Pro Leu Arg Tyr Glu Thr Leu
125      130      135      140      145      150
Met Asn Gly His Val Cys Ser Gln Leu Val Leu Ala Ser Trp Leu
140      145      150      155      160      165
Ala Cly Phe Leu Pyp Val Leu Cys Pro Thr Val Leu Met Ala Ser
155      160      165      170      175      180
Leu Pro Phe Cys Gly Pro Asn Gly Ile Asp His Phe Phe Arg Asp
170      175      180      185      190      195
Scr Trp Pro Leu Leu Arg Leu Ser Cys Gly Asp Thr His Leu Leu
185      190      195      200      205      210
Lys Leu Val Ala Phe Met Leu Ser Thr Leu Val Leu Leu Gly Ser
200      205      210      215      220      225
Leu Ala Leu Thr Ser Val Ser Tyr Ala Cys Ile Leu Ala Thr Val
215      220      225      230      235      240
Leu Arg Ala Pro Thr Ala Ala Glu Arg Arg Lys Ala Phe Ser Thr
230      235      240      245      250      255
Cys Ala Ser His Leu Thr Val Val Val Ile Ile Tyr Gly Ser Ser
245      250      255      260      265      270
Ile Phe Leu Tyr Ile Arg Met Ser Glu Ala Gln Ser Lys Leu Leu
260      265      270      275      280      285
Asn Lys Gly Ala Ser Val Leu Ser Cys Ile Ile Thr Pro Leu Leu
275      280      285      290      295      300
Asn Pro Phe Ile Phe Thr Leu Arg Asn Asp Lys Val Gln Gln Ala
290      295      300      305      310      315
Leu Arg Glu Ala Leu Gly Trp Pro Arg Leu Thr Ala Val Met Lys
305      310      315      320
Leu Arg Val Thr Ser Gln Arg Lys
320

```

<210> 21
<211> 318
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472457CD1

```

<400> 21
Met Asn Pro Ala Asn His Ser Gln Val Ala Gly Phe Val Leu Leu
1      5      10      15

```

WO 01/66742

PC/US01/06814

Gly Leu Ser Gln Val Trp Glu Leu Arg Phe Val Phe Phe Thr Val
 20 25 30
 Phe Ser Ala Val Tyr Phe Met Thr Val Val Gly Asn Leu Leu Ile
 35 40 45
 Val Val Ile Val Thr Ser Asp Pro His Leu His Thr Thr Met Tyr
 50 55 60
 Phe Leu Leu Gly Asn Leu Ser Phe Leu Asp Phe Cys Tyr Ser Ser
 65 70 75
 Ile Thr Ala Pro Arg Met Leu Val Asp Leu Leu Ser Gly Asn Pro
 80 85 90
 Thr Ile Ser Phe Gly Gly Cys Leu Thr Gln Leu Phe Phe Phe His
 95 100 105
 Phe Ile Gly Gly Ile Lys Ile Phe Leu Leu Thr Val Met Ala Tyr
 110 115 120
 Asp Arg Tyr Ile Ala Ile Ser Gln Pro Leu His Tyr Thr Leu Ile
 125 130 135
 Met Asn Gln Thr Val Cys Ala Leu Leu Met Ala Ala Ser Trp Val
 140 145 150
 Gly Gly Phe Ile His Ser Ile Val Gln Ile Ala Leu Thr Ile Gln
 155 160 165
 Leu Pro Phe Cys Gly Pro Asp Lys Leu Asp Asn Phe Tyr Cys Asp
 170 175 180
 Val Pro Gln Leu Ile Lys Leu Ala Cys Thr Asp Thr Phe Val Leu
 185 190 195
 Glu Leu Leu Met Val Ser Asn Asn Gly Leu Val Thr Leu Met Cys
 200 205 210
 Phe Leu Val Leu Leu Gly Ser Tyr Thr Ala Leu Leu Val Met Leu
 215 220 225
 Arg Ser His Ser Arg Glu Gly Arg Ser Lys Ala Leu Ser Thr Cys
 230 235 240
 Ala Ser His Ile Ala Val Val Thr Leu Ile Phe Val Pro Cys Ile
 245 250 255
 Tyr Val Tyr Thr Arg Pro Phe Arg Thr Phe Pro Met Asp Lys Ala
 260 265 270
 Val Ser Val Leu Tyr Thr Ile Val Thr Pro Met Leu Asn Pro Ala
 275 280 285
 Ile Tyr Thr Leu Arg Asn Lys Glu Val Ile Met Ala Met Lys Tyr
 290 295 300
 Leu Trp Arg Arg Cys Phe Asp Pro Ile Gly Pro Leu Glu His Arg
 305 310 315
 Pro Leu His

<210> 22
 <211> 1348
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> misc_feature
 <223> Incyte ID No: 536482CB1

<400> 22
 gaagccag tccgcccgg cttccgggga cgcacagggc aggcgctggg catgcccag 60
 ctccgagcc gggcagctg gagtctccag gcgctgcaat ctgttcgcc gcccccggc 120
 ggagccggca gatccgggto ttgtggttaa gactgacgtg gtaactcgaa tcaaaaacaga 180
 ggagggggag gaagccggcg gccagaaacg gcagtgccag cagcgtccgg agcagccgca 240
 gccttctgga agtcccaggc ggtctttctg ccagccctcg gtcccggccc ccactctccc 300
 gcccacatg gttgtgtct gggcggattt aaneogtoaa gtaaaatcaa gctgggtat 360
 catggcaga gttgatttg atcccggga atgtgttgc tctcatgac atgcaatgag 420
 aagctgato atctgttac ggcagccca gtctactgc accgacacg agtgtttaa 480
 ggaattacg ggaacctctg gtgtataatg catcagttt acaatgatct tggtagcct 540
 gatggtatt ccattgatct tgttcttaact gagacctct aatctaagag gatccagcct 600
 acgtgaaag caaccagtc ctcataatg acaagatcca ccagctctc ctgtgacta 660
 accttggat atgggaagt aaataagtta acaccttgc ccaccaaac aacyaagatg 720
 accagatca ccttaacccc attagaactg tttttctct tttatctgca atatgggatg 780
 gtattgtttt catgagcttc tagaaaattc acttgcaggt ttatttttgc ttccctgttt 840

WO 01/66742

PC/US01/06814

```

atctgcattc ctatttacag tataattgag tgaatgaba tatttttaa aagttacalg 900
gggttttttt ggttgcccta aacttacaaa cttcccaact attctgtttg taactgtgat 960
tataattttt gtagaattt ctggcctgat tgaaggaaat ttgtaggttc tgcatttata 1020
tattttaaat agatttgata ggttttttaa ttgctttttt tcataagpta tttataaagt 1080
tatttggggt tgtctgggat tgtgtgaaag aaaattagaa ccacgctgia tttacattta 1140
ccttggtagt tatttgggg atggcagttt cctgtagttt tggggactgt ggttagcttt 1200
ggattgtttt gaaattttaa gctgaaactt gttctctgga ttaaactgdc tbatgtggct 1260
agaaatggaa gagagaaaaa atgaaatggt tgtttactaa ttttactc ccattaaaaa 1320
tctctatgt taagaaacc ttaataaa 1348

```

```

<210> 23
<211> 1446
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<222> Incyte ID No: L316020CB1

```

```

<400> 23
gaaagccacc agtctcaag tgaacattaa gtgagatgat tctagttaac gaactagaac 66
aaatlccagc acatagttaa atzicaggga aaltctgta ctgttatgt tgggtgagct 120
gaactggatg tagatgtttt cctctctctt gctgacccct ccgcagttt tgtcttgtga 180
tgccatcaac acatctctcc cttctgacc tggctctgce ccattgtgtc ccaagaaat 240
cgtgagaata gttagccccc cgtctcccca cctctgttgc tctctgtgta gttgttcaaa 300
tgagttgaga agttgagag cttttgccta ttgaagytgc actgagaata aactctttcc 360
tgccaccaga attgcaatgg ttcacggcct gcaactcatt ccatgaaac agttatbagc 420
caacgaatg tcaactaaag caaagcagcc agggctcact cgtgttgaga ctgagctcc 480
tcagcccttg gatctctccc ctggtctctt tgaagctcag tttctcatt ggttaaaagg 540
aagtgaaaga gttcttcaaa ggtctactac agagatbaaa tgaataaaat gaatacaact 600
agcagaggag ggtgtggtgt ttaaaagtca cagatggggc cccctggggc cctccggccc 660
agtgtttctc ttgaccctca tgaatctcat tttttgttac atcccattag tgqcccatat 720
ttaaaalgt ggagatttca cataaaatta aaagytctgc atttctttt ttcttttctt 780
tttttttttt ttgagacaca gttcactctt gttccagggc tagagtgcag ttggcagatc 840
tcagctcact gcaacctctg cctcccaggt tcaagtaatt ctctctctc agctcccaa 900
gtagctggga ctacagggac gtgcccacc gccccagctaa ctlltgtatl tttagagag 960
atgggtcttc accacatgg ccaggttgg ctgactccc acctctgat ccccccagct 1020
cgtctccca aagcctggg alllacaggcg tggccaccg cgcacagca agtctgcat 1080
tttcttttag aactcagaac acccaatgct actagcccac aatctcgca ttgcaagaa 1140
ctaaaataag acctctccac tgcaggttgg ggcattgccc agcctatggt ttgccactc 1200
cctctctttt ctccgttttt tcatlaattg tgaactgac ttgcalcacc ctbtctgtc 1260
agtctctccc aaactctctt gcttgccccc ctctagtga atatattgt gottaccoca 1320
atatactgt gtagctattg aactclatbc gtgactgct tgaactaat tcatlgtcat 1380
cataaatat tcatatccaa taaactattt aaaaagtga gataagaaac cyaaaaaaaa 1440
aaaaaa 1446

```

```

<210> 24
<211> 1463
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<222> Incyte ID No: 2816437CB1

```

```

<400> 24
cagatgggtc tctgctcda gactcccgaa tggctgggac tacaggoatg tgcaccatg 60
cttggctaatt tttgtattt ttatagggga tggctgtcca ccatgttggc aaggatggtc 120
tggttctctt gacctcatga tccgcccacc tgggattact tatatgaaaa taaaatttta 180
aataaaaaat agcatctgat tacactatb ggtgagacta ttagtgtgaa gtcatacttt 240
tacltcaatt gacaaaaaaa cactctctga tatgggat ttgactctat tgcacaaaa 300
gocatacaaa tattctgatt tagaatcaaa ctctctctc gctgtatatt tgcagcttt 360
tataaatat tacgggagct caatgaaat caacaalat aatcttlat taccacaac 420
atlatgatg cttatgcttg tttttctaa antctatag ctlltlatg cacatttata 480
cttttctcag tttttatc tgaagaatgg ctgtctctgc aaaaaactgc tattctcagc 540
tgtatccaca atgactaata gtgagtgga gtgagttgga taaaagcaaa tgtgtctct 600
ctgattttt tttcatcaat tggctaaaa acagggagat gcaagagatg aagaaaaa 660

```

WO 01/66742

PC/T/US01/06814

```

caatccctga aagatcatga agaagctctc aatcagacaa aaaaaccatc agggagatbt 720
tttttttttt cattattgta ctctctgctt ctatgagllt aaalttcltt tacacttcc 780
atagaataaa aortaggctt acatagaag aaacctagcc ttttggccag gaatggtygc 840
ttaagctctt aaltcccgca ctlltggagg cuuaggcyag cggatcatct gaaglcagga 900
gttcaagcaa atcagggggc ccgtgaaagg caactgagca gggatccatg ggaagacc 960
ctcaaaagc acaagattct ctgttaact ttcaatttgc tgatactta gttaaact 1020
cotgggaac gtctgcatta gttctctct ctgtagttct tcttaccctg cctgtaaac 1080
aaaacctatc tagtctctgc ataggtttcc acttcttgc ccaacctgag gaatggaag 1140
caacggcaca gtccttgcLc atgttttggg gtgaaggag ctlyaaagtc atgtgactt 1200
tycaaggct tctctctggc tcatgtcaga tacagctct aactcccaag cagctacca 1260
tagtctctc ctttttttgc gtgtgtgat gggtttgcga ctgtttgenc agctggagt 1320
gcaatcgtta caplctctgc Lcaactgcaac ctccgctcc caggttcaag tgaatctct 1380
gctcagctct ctcaagtaaa tyygattaca ggcalgggc actaagggac gtagactct 1440
ctcatatgy tottatggcc aat 1463

```

```

<210> 25
<211> 1435
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 2289894CB1

```

```

<400> 25
caaacatcaga gtcacagcca gtaccatctc cgacaggggc tgagttgctg gagctgggct 60
ggggcagggg agaaagacag cactcactct tgcacacctc catgggctg gccaaagccc 120
cgaagagatg gcagcttggg cgtctggagcc acctctctgg cagccaatga ggtgaggggc 180
cggagagcga aggacacaga ggaactagag acagtgatg gaactctcgc agclttaggc 240
tccattctgc catctacatc ctagggcagg gttaggctct agagcccaaa tggccaactc 300
cacaggctct aacgctctcag aagtcgcagc ctcttggggc ttgatctctg cagctctctg 360
ggaggtgggg gcactgctgg gcaavggcgc gctctggctc gttgtgctgc gcaagcgggg 420
actcggcgac gegtcthacc tggcgcacct gtgctctggt gaactgctgg aggcggctc 480
catcatgctg ctgggctctg tggcggcacc gccgcccggc ctgggcccgc tggccttggg 540
cccgggca tyggggcgg ctctctctct ctccgcctct ctgtctcgg cctgcaactc 600
cgggtggcc gcacttggcc tggcaagcra ccgactcact gttgcccgc tggggcagg 660
ctcggggcgc ccgclatgc tctgtctcac ccctgttgg gccggcggc gaactctggg 720
cggctctctc ctgctggccc agcagcccgc acctctggcc gctcctgctc gctgtctgt 780
ctggttggg ggcctggggc cctctcggcc gctctggccc ctgctggccc tccgctgccc 840
gctctctctg ctgctggggc cctacggcgg catctctctg gttggcctc gctctgccc 900
gagcccacca cggcggggc gggggtcccg actcctctg gaactctctg atagcggct 960
ttccalcttg ccgcccctcc ggcctgccc gcccccggcc aaggggccc tggcccagc 1020
gctggcctg ggcacatcty cagcctgcty gctgctctat ggttggctt gctgtggcc 1080
cggagggg cccggggag ccagagggc tctcacttgg gctgctact aggtcttgc 1140
ggctcccccc ltcctgtacc ggtctgtgca ggcctcctg cgtttgcaac tggcggct 1200
ctctctggct gaactgctg gaactgtgct ggcctgcaact ccgcaagctt ggcacccggc 1260
ggcaactctg caatgctcc agagaccctc agagggcct gccgtaggcc cttaaggagg 1320
tccagacag acccccaggt tggcggagg gggggccc gcataccagg ggcacctct 1380
gagttctctc tctgagcag gagnaaggag ggtggtttcc gttggggctc atcca 1435

```

```

<210> 26
<211> 2147
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7066050CB1

```

```

<400> 26
ctcggttca ggcagcgcga ctgcccgtgg cgcacagaca gggcgcagac cttggggccc 60
gggcccacag ggtctggggc tctctgggccc ggcggcggct agcaggtca tctgtctgca 120
ttacaactac accgcaagc tccgggttgc cgcctaccag ccggctgccc gctctgccc 180
cgaagcggct gttgtgctgg ccgtgtgctc ctctactctg cttagaatac tagctgtgtt 240
gttgggtctc ggaagccacc cggggttcca cgtccccat tctctgctcc tggcggcct 300
cagcttctc gatctgctg cagggcccgc ctacggccc aacatctcac tgtggggccc 360
gctcagctg aaactgtccc ccgctctcty gttcgcagg gggggaggc tctctgtggc 420

```

WO 01/66742

PC/T/US01/06814

```

actcaactgg tccgtgctga gccctcctgc cctccgctg gagcgcagcc taccatggc 480
gagagagggg ccagcgcctg tctccagtcg gggggcagc ctggcagatg cagccgaggg 540
ctggggctgt tccgtgctcc tccggctcct gccagcctg gcttggatt gccctggctg 600
cttggagcgt tcttccactg tcltgcctgt clacgccaag gcttaactgc tcttctggct 660
gctggccttc gttggcctcc tggccgctat ctgtgcaacc taaccggcca tctactgcca 720
ggtaaccgoc aatggcggc gcttaccgag acggccggg aclyggggaa ccaactggac 780
ccggggcctg ccagcgcctg gctgctgctg ctgtgtctgc acgctccagg tggctgctcc 840
ggccttctg gcatgttggg ccccccctt cctgtctcty llyctcagc tggctgccc 900
ggcgcagacc tgtcctgtac tccctggcgc ccatccctc ctgggactgt ccatggccaa 960
ctcactctct aaccctatca tctaacagct caccacccg gactgggoc accgctcct 1020
gogcctgttc tcttggggac gccactcctg cggcagagac ccgagtgct cccagcagtc 1080
ggcggagcgc gctgagggct ccgggggccc gggcgcctgc nlgcccggg gccctggatg 1140
ggacttccag ggctcggagg gctcactgccc ccagcggacc gggctggaca ccagcggctc 1200
caccggcagc ccgggtgca cccagcggc ccggactctg ctalccaaac cggctcagca 1260
ctgacacctt ggcaccagca clgtctctcc agtlttacc gactgtctt ttttaactaa 1320
aggaatttgt aggaatgca gccaaagtg cagtgccaaa agatgcagg gaaatgtatt 1380
tatgcagcga caccaccaca tglgaacaaa cagacaaaa atctgtccc tctggaaatt 1440
gacttctctc ttgggaacac agaaaagaac tgggtgatga aataatggag atgattccag 1500
tgacaacaga cagagatggt gatgttggct agggagaco tctctgcaga ggtatgtact 1560
tgtgatgta gctgagacc ctgtcctggg aagacaaaa gaaaagcatt tcaaggtgag 1620
ggacttggct ggcacaagg cctgagcctg aactgtccc atgtgtctc agaactgac 1680
cgttgttgtt tggccgggg aggtgcaat gggagagac gggagagac agagctgca 1740
aggatagtt ccccaaggac ctgtgggtg atatagagg ctctgcttt cctctgagtg 1800
aggtgggagc catagaagct tctaagcaga agagggactt gccctaatto aggtgctac 1860
aggttctctg tggcctccab gggaggttga aaaccagaya aggtgaaagg gggctgact 1920
gagccacagg acaaatgaty gagatccag cttagccag acccctgga tctagatag 1980
atcttagagg aagcagaaag natfactgag gaattgagtg taagatgga ataaagtat 2040
caaggcaatt gccagggtg gggcaccccc aatllgact ctgggagact cagcccaatc 2100
ctactgtgta ataaaatttc ttttttttt ttaaaaaaa aaaaaaa 2147

```

- <210> 27
- <211> 1989
- <212> DNA
- <213> Homo sapiens
- <220>
- <221> misc_feature
- <222> Incyte ID No: 5376785CB1

```

<400> 27
gectcaqcaa tggcggagcc tectccccca gccctcctgc cgtgactctg attttccagt 60
gaggacattg tggctcctct ctgctgcca gacttccctg agcctagctga cctctctctg 120
aagtaacctg acttggccac ctctcaactg ctctcaactc tggcccctct cactatctct 180
gtggctctag ctgtgtggc caagaaac tggctgtgta atatgtctg caggtgacc 240
acagagcagt acttggcctt gggcgcanaa agagaagaga ccaataagat gttgatctg 300
gtgtgtacc tcltggcct ctgtgtgttc cccctcaact gctagctctc cctctctctc 360
agcaaggcca tccgcaccua caatgccctc tactttgctc tccactggtt tggcctgagc 420
agcaactgct ataacccctt catatactgc tggctgaag agaactccag gattgagca 480
aagccttacc tggagctgtg tcaaaagact cccaagcctc agsgggcag gcaaccctcc 540
ccgttccctt ccttcaaggtt ggcctggaca gagaagaatg atggccagag ggtctccctt 600
gccaaacacc tcttggccac ctcccaactc cagtctggga agacagacct gtcctctgtg 660
gaacctctg tggcgttagg ttagaagagg ttggcaagag ggaagggag ggtctctgtt 720
ccacttggag cagggaagaa ggccttcttc tcaacatgta ccttcaagat gctggcaaca 780
cactcttcca gaagctgtag gactcttgan ttcttaggaa actgtccagg cctctagccc 840
cactgtgagt gaaaactaaa aggcaccacc aactagacat gtttctabaa atctccactt 900
aagaaacact gggagggcaca gcagcctgta tctctggaga agaggagcga ggcacacgct 960
ggcccaagat gggctgcaat catccaactg cctccactct gggggcagct gctgcttacc 1020
agccttctct actgtctgagc atccgcaagg gagacctaaa tcaactctt gglgtgtgca 1080
cccagatgca tagagctctg ctggaacag alacacggc cagggaactg ccagcaagcc 1140
agcctggggg tggagcttt talgctcac ttcttggagt cactggcca tggatgata 1200
taagtcttca gggctctagc aatatccaga taagaagaga ccaactggg tctcttanaa 1260
caaggggaaa tttatttgc caacttagaaa aalttagaaa agccacact cactatccca 1320
cacacaaaat cactctctta tcccctccat ttgtgataac atctgtgac atgtctggc 1380
ctatatttca acatctctct tctgtgtgtt gattgtgtgc atgtgtgact gaccttttt 1440
ttttctttt ttttttttg gagaatgga aattctctg tctgtctcc cccaggttg 1500
tgggttctc ggttggccc aactctctg cctcaactg aaagctctg cctcccccag 1560
gtttcagatg cgttctctc ctgtgctctc agccctccc ttaggttaag ctggggcacl 1620

```


WO 01/66742

PC/T/US01/06814

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472363CB1

<400> 30
atgaticatg gngyagatcc aaacatcaac attaacaggc gtttggpaga agcccattcc 60
aacctcaatg ataacgcttga ggggttccagc acttcaatgg aggaagcaac tgcagatag 120
gtggaantag caagcgaaat gcanttagaa gtgaaactg aagaaggcaac tgaatgctc 180
aetctacga caaaaggctc ggaagacttc cacatgtgga tcccgggccc ttctgctct 240
gtttaccctg tggctttgct gggcaatgcc accattctgc tagtcaaca ggtagaacag 300
actctccggg agcccagctt caacttctct gccattcttt ccaactatga ttggcccct 360
cttgcaccc cttgtccctg caatgtgggt atctctgggt ttgatgctca ctagattaac 420
tatggagctt gtgtggccca gactttctct atocctgccc tcaactggca gtaggcttag 480
gtcttaatgg ctatggcttt ggaacgttat gtggcactc ggtctcact acattacga 540
accatttga cacccttagt gttggggggc attagcactg gcattgtaat tctccctgt 600
tcaattcac tcccatggt ctatcttacc taccgctcac cctttgtca ggtccacta 660
atagcccatt cctactgtga gaacatgggc attgcaaat tgcctgtgg aanaattgt 720
atcaatgta tctatggct tttttagtct tctttcttg tctgacctt ggtgctcatt 780
ggcactctgt atgtttacat tctccgtgct gtcttccgpc tcccatcaca tgaatccag 840
ctaaagccc taagcagctg tggcgtctat gttggagtea tctgtgttt ctatctcct 900
tcagctctct ctctcttacc tcaatgattt ggaacccaaa taccaggtta cactcaact 960
ctgtgtgca atctctatct gattacccc cccctctca acccaactc ttatggggt 1020
aggaccaac agattcgaga gtagtgctc tatgttttta ctaaaaaata a 1071

```

```

<210> 31
<211> 1001
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472364CB1

<400> 31
actgtgattt ggaaaaaatg tttatcaca caagagcata ttcccccag tencatttt 60
cctaatgga atcccagctc tggagactt ccaactgtgg atctccggg attctctctc 120
tgtttacctt gggctttgca tggcaatgc ccaactcttg ctatgctca aggtgagca 180
gactctcggg gagccatgt tcaactctc ggcactctc tccattattt atttggccc 240
ttctgcaacc tctgtccctc ccaatgtggg tatctctcgg ttgatgctc acagattaa 300
ctatggagct tgtgtggccc agatgttctc gatccatgcc ttcactgga tggagctga 360
ggcttactg gctatggctt ttgaccttca tgtggcactc tgtgctcaca taactaacg 420
aacactctg acactcctag tgttgglygg catlaacatg tgcattgta ttctccctt 480
tttaattaca ctctccctgg tctatcttat ctacccgcta cctctttgto aggtctcact 540
aaagaccat tccactgtg agcaatggg cactgcaca tctctctgag gaacattgt 600
tatcaatggt atctatggg tttttagt ttctttctt gttctgacc tggctctct 660
tggactctg tagttttaca tctccctgca tgtttccgc ctccctcacc atgctgctca 720
gctaaaaagg ctaaacagct gtggcgctca tgttggagtc atctgtgtt tctatctcc 780
ttcagctctc tctttctca ctatctgatt tggacaccaa ataccaggtt acattcaat 840
tctgtgtccc aatctctat tgaattatcc accctctctc aaccccata tttatgggt 900
ggagaccaaa cagattcgag agcaggtgct ctatgtlct actaaaaaat aagactctta 960
ccatgttat ttaactaagg cttgactct tctataaaga c
1001

```

```

<210> 32
<211> 1065
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472434CB1

<400> 32
atgaggctcc tgaagcagc ggcgaagcag actgctatg tggcagggga gcaagaggca 60
ggaatacctg agcttggct tloccaggga gaagttagag cagctctgca tggcagtgga 120
ggtaaccctg gtagaccac agcctcggcc accgctcctc ttgcaaaact ggtcccaact 180
gaccgacct caccagatt cgtgctctgc tctctctctc gttgacatc attgtcagta 240
tcattctcc tacaagcaa tatggagccc agaaaaacc tctctctcat ggacatctg 300

```

WO 01/66742

PC/T/US01/06814

```

ggccctccct cctttgtgcc t.tctatgcta gacaatltno lggaaaccca gaggaaccatt 360
cccttaccctg gctgtccctt cagatgttac ctgcccctgy cgtgggctt aaaggagtgc 420
ctgctgctgg ctgtgatggc abatgaccgr tatgtggcta tctgccagcc gclltagtac 480
ccagagctca tgagtgggca gacctgcatg cagatggcag cgtgagctg ggggacagcc 540
tttcccactt cactgtataa gtcacacctt gctcggcacc tcccctctg tgcccacgtc 600
atcaactact tctatgagat ctgtggcagtg ctaaaactgg cctgttggga cactcccttc 660
aatgctgtgg caftaatagt gcccacagcc gctctgaacc tggcccctt ctgtctaatc 720
tgctctgttt accttttcat cctgtctgcc alctttaggg tccccttgc tgcagycgg 780
tgcacagcct tctccacctg ctccagccac cgcacagctg tggtygtttt tttgtggaca 840
atctctctca tgtacttcaa acccaagccc aaggatccca acgtggalaa gactgtcgca 900
tggttctaac gggttgtgac gccctcgtct aaccccatac tttaacagct gaggaaatgca 960
gaggtgaaag ctgcccctct aactctgctg agaggaggtt tgcctctcag gaaagcctcc 1020
cactgtactt gctgcccctt ccccctgtca gctggcctag gctag 1080

```

```

<210> 33
<211> 963
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<222> Incyte ID No: 7472435CB1

```

```

<400> 33
atggaaaag ccaatgagac ctcccctgtg atgggpltcg ttctctgag gctctctgcc 60
cacccagagc tggaaaagac attctctctg ctcatctctg tgaatgact cgtgatctg 120
ctgggcaatg gggctctcat cctgggtgacc atctctgact cccgctctca cagccccttg 180
taattcttcc tagggaaacc ctctctctctg gacatctctg tcaactacc ctcagctcca 240
ctgtctctgg acaccltctt gctccccag gaaaccactt cctctctcagc ctgtgtctg 300
cagatggcag tctctctctg ctgtggcagg acagagctct tgcctctggg cctgatgca 360
tttgatgctc atgtggcact ctgcaacccc cttaggtact cctgtgactf gaggaggtct 420
gctcaatgc ccatgctgc cagctctctg gctatctgct gctgctcttc cgtggtacac 480
acatctctgg caatctagct gccctctctg ggagacaatg tcatcaacca ctccacctgt 540
gagatctctg ctgtctcaaa gttggcctgt gctgacattt ccatcaatgt gatcagctg 600
gaggtgacga atgtgatctt cctagagctc cgggtctctg tcaatctctt ctctatgct 660
tcatcaatca ccccactctt gaggatctcc tcaagctgag ggagyaaaa gctctctctc 720
acctgctctg ccccctctcc cgtgtgctct gctctctcag ggcctctctt ctctatgct 780
gggaaagcct agtctcaagc ctcaatggga gtagacaatg aggatctctt agacaactc 840
atcccccttt tctatggggt ggtgaccccg atgctcaacc ccatcaatca tgcctgagg 900
aaccaagatg tgaaggctgc tgtgaggaga ctgctgagac caaaaggctt cactcagtca 960
tgg 963

```

```

<210> 34
<211> 1101
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<222> Incyte ID No: 7472438CB1

```

```

<220>
<221> unsure
<222> 545, 547
<223> a, t, c, g, or other

```

```

<400> 34
atgaaactct aaaaagaacc aaccaaaact ccttacaaga ataaactyga agyaataaaa 60
ctcaactcaa cagaaataga aacatttagg acanagana gaaaccaatc agataaagc 120
agaaagggg cagaaacagc agatctcagg tggctctccc ccagagggg aatggctgca 180
ggaaatctct ctacagtgac agagtcaatt ctcaaggtt caacagagag agaaagccc 240
cagctcccc cctttctctt ctctctctgg atctactctg tcaactctgt gggaaactg 300
ggcatgatca ctctaatitg tctgaactct cagctgcaaa ccccctgta ctactctc 360
agaaatctgt cactcaatgga tctctgctac tctctcgtca ttaccctcna gatgtggtg 420
aaettctgtt cagagaaaaa cactctctcc tctcaggggt gcatgtcaca gctctactc 480
tctctgttt tigtcaatgc tgagtgttac atgctgacag tgatggcta cyacagctat 540
gttgnctct gccacccttt gctttcaaac atcattatgt ctctcaacac ctgctgtgt 600

```

WO 01/66742

PC/T/US01/06814

```

cgggtggctg tggletacgc catcggaactc attggctcca caatagaaac tggcctcwtg 660
ttaaactcgc cctattgtga gcaactcactc actcactactc tctgtgaact cctccctctc 720
atgaagclgt cctgctctag caactatgat gttgagatga cagctctctt ttcggclgga 780
ttcaactcca tggtaacgag ctttaacagt tttgtttctt acacctctat tctctccagg 840
atcctcggca tcagcaccac aggggggaga tccaaagcct tcagcaccg cagctccacc 900
cttgagcag tgggaatgtt ctatgagatc actcpactcw tgaacttaaa acctctacw 960
atcagttctt tgaaccggga gaatgtggcc tctgtgtrct acaccccggt aatccccagt 1020
ttgaatccc taactcaacg cctgaggaae aagguagtaa agctgcctt gcagaaacy 1080
ctgaggggta aactgttttg a

```

```

<210> 35
<211> 933
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472439CB1

```

```

<400> 35
atggagcaca aaaactcttc tgtgacagag tttatctctg aaggcttaac ccaccagccg 66
ggctctcaga tcaactcttt ctctctgttt ctgggtttct acarggtcac cctggtgggg 120
aaactgggct tgaLuacct gatgggctg aactctcaac tgcacactc calgtacttc 180
tcccttttta acctctcttt atagatcttc tgtttctcca ctaccctac tcccaaaatg 240
ctgagagttt ttgtctcaag gaagaacatc alttctctca cagggtglat gactcagctc 300
ttctctctct gctctctctt gctctctctg tctctctctc tctctctctc tctctctctc 360
cgtcagctgg cctctctctc cctctctctg tctctctctc cctctctctg cctctctctg 420
ttgtctcttt tgtgggtggt ctatggggtg ggggtctctg gggcctctgc ccacacagga 480
agctcaatga acctgacttt ctgtgtgac aacctgtctc actctctctc gttgtgactc 540
ctctctctct tctctctctc ctgcaacagc tctctctctc atgagctggt ggtctctctc 600
gtgtgtgtgt tgaagctgg aaagccctat gtcactgtct tttctctctc tgcctctctc 660
ctctctctca tctctctctc cagttctctc gaagcaggtt ccaagcctt tagtactctg 720
agttctctca taatgtgagt tctctctctc ttgtgtctg gtgtctctc gtctctctca 780
cctctctctc cctctctctc ctgagcaggg aaagtctctc cctgtctctc taactctctc 840
gtccctgtct taaccctctc aactctctag ttgaggaaac aggatctctc agttctctct 900
aggagacttt tgggcagaaa aatctctctc taa

```

```

<210> 36
<211> 936
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472440CB1

```

```

<400> 36
atggctgctg agaattcttc cttctgacac cagtttatcc tctcaggtct aactgaccaa 60
cgggggtccc agatccccct ctctctctct tttctaggtc tctcaggtct cactgtggtg 120
gggaacctgg gcttgataac cctgataagg ctcaactctc actgtcactc cctctatgac 180
tctctctctc ataacctctc cctctatgat ctctctctct ccagtgctat cactcccaaa 240
atgtctatga gctgtgtctt aaagaagaac agcctctctc accgaggtg tatgactcag 300
ctctctctct tctctctctc tctctctctc gctctctctc tctctctctc aatggggtat 360
gaccgtatg tggcactctg tcaaccactg ttgtactctg tcaactctc tcccaaggtg 420
tctctctctc tttgtgtgg tctctatagg atgggttttg ctggggcact ggcaccaca 480
gctgtctatga tgggtgtgac cttctgtctc aatacctctg tcaaccactc catgtgtgac 540
atctctctcc tctctgagtg tctctgaccc agcactctat tgaatgagct tctgtgtgtt 600
gtctgtgtgg gacttgatct tgggtgtctc acagtcaaca tctctctctc ctatgtctc 660
atctctctca gactctctca cactgtctct ccgggggaca ggtccaaagc ctctgacacc 720
tgaagctctc aataactctc agttctctct ctctctctct ctggagact catgtactc 780
aaactctctc ctctctctct taktgaaccg ggcactgtat ctctctctct ctctctctct 840
gtgtgtctca tctctcaacc attaactctc agcctgagga ataaggactt caaagtgtct 900
ctaaagaaaa tcttgaaaa aaatgactct cctctca

```

```

<210> 37
<211> 945
<212> DNA

```

WO 01/66742

PC/US01/06814

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472443CB1

<400> 37

```

aTggooaaga ataactocac cagagtaacc gaattcattc toatgggctt tatggacoac 60
ccccaaatgg agatbecect ctttctgggt tttctagatt tctactagt ccccccttt 120
gggaatgtgg ggratattat gtttaatocaa gtatgatgta aactctacac ccccaatgac 180
ttcttctctga gccacccctc cctgctggat gccctgtfaca cctcagtcac cccccccag 240
atcctagcca catlggcccac aggoaaaaag gtoatctctc acggccactg tctgcccag 300
ttctllllat toaccatctg tgcaggcaca gagtcttllc tggctggcagt gattggcctat 360
gatctgtatg ctgcaattcg caacccactg ctctataccg tggccatgaa tcccaggctc 420
tctggaacc tggtagggag agctatgtlc tggggggtg caggagccat cctgggtaac 480
actggaacct taccctctc ctctgttaag gacaaTcaaa kaaactctt ctctggtagc 540
ctccccccc tgotgaaget tgcctgcaat gacaagcaaa anactgagat tgtcatcacc 600
tctctggca allttgtgat ttggccaat gccctcgta tctgtattc ctatctgctc 660
atcatcaaga ccattttgaa agtgaagctc tcagggtgga gggccaagac ttctccaca 720
tgtgctctc acatcactgc tctggccctt ttctttggag cccctatctt catgtatctg 780
caaaagggct caggcaaatc tctggaggaa gacaaagctg tgtctgtctt ctatacagtg 840
gccatcccca tgotgaacc tctgattcac agcttaagaa caaaagatgt aaaagagcvc 900
ttcaaaaag tgottaggag actccagggt tccctgagca tctag 945

```

<210> 38

<211> 1041

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472445CB1

<400> 38

```

atgaatgcag cttaaagcca ttatcctaag cgaactaatg caggaacagg aaaccaata 60
tcaaatgttc ttaactgmaa acaggcnaaa atataaatgg gagaagnaaa ccaaaccttt 120
gtgtccaagt taaactctct ggtctllcca caggacttgc agaccagat cctgctatt 180
aacctcttcc taaactctta tctgtgmoa tctgttggac accagctcat cctactctc 240
atciltcttg aittctgctc tcaactctcc atgtatittt ttcttagaaa tctctcttt 300
gcaatctctc gttctctcac tagcatttc cctcaagtgt tggttccell ctgggtaag 360
agaaaacca ttctcttcta tgggtgatg acaagataa ttgtctctc tctggtggg 420
tctacagagt gtgctctgct ggcaytgatg tctatgacc ggtatgtgc tctctgcaag 480
cccctgtact actctaccat calgacacaa cgggtgtgct tctggctgta ctctaggctc 540
gggcccagtg ggcactagt gtcttactta gatcaagct ttacttcca tcttccctac 600
tgggacaga atataactaa tcaactctll tgggaactc ctgcccctc gaagctggct 660
tcaatgaaa cttaacgcaac agaaatggcc atctttcwa tgggctgggt aatctctctg 720
gcccctgtct cccgattcl tggttctbat tggaaatata tctccactgt tctccagatg 780
cagctctggg aayggagact caaagctttt tcaactgtg ctcccactc tattgllgtt 840
gctctctct atgggtcag aatattcacc taactgagc caaactocaa gactacaaaa 900
gaactggata aaagatata tctgtctctat acagcgtgac ctccaatgtt gaaccncaa 960
attatagctc taaggacaaa agatgtcaaa ggggctctca ggaactagt tgggagaaag 1020
tgcctctctc ataggcagtg a 1041

```

<210> 39

<211> 951

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472446CB1

<400> 39

```

atggagctct ggaacttcac cttgggaggt ggccttcatt tggtagggat tctgactgac 60
agtaggtctc ctgaaactgct ctgtgtctaca attanaatcc tatacttggt ggcctgac 120
agcaatggcc tactgtctct ggcctatcacc atggaagccc ggcctccact gccoatgac 180
ctctgtcttg ggcagctctc tctcatggac ctctgttca catctgttgt cactcccaag 240

```

WO 01/66742

PC/T/US01/06814

```

gcccttggcg aettttctgg cagagaaaac aacacatcct ttggaggctg tgccttcag 300
atgttctctgg cactgacaat ggttggtyct gagggcctcc taactgctct catygcctat 360
gacagatgat tggccatbtg taatcctctg acatacatga cctctatgag ctcaagagcc 420
tgcctgctca tggctggccc ctctctgctc ctggcatccc taagtgcctt aalatatacc 480
gtgtataaca tgcactatcc ctctctgagg gccccaggaga tcaaggctct tctctgttag 540
atcccacaat tgcctgaagg ggcctgtgct gatacctcca gatctgagct cactgtatct 600
gtgatgggtg tgaactctct gattccctct ctgtgtgcta taactggctc ctatacaca 660
actctactca ctgtgtctca ttgtccatca aatgagggga ggaagaaagc ccttgtcacc 720
tgcctctccc actgactgt ggttgggatg ttctatgag ctgccacat catgtatctc 780
ttgccagttt ccttccacag caccagcaca gacacatca tctctgtttt ctacacaatt 840
gtcactccag cctgtaatcc actcactcac agcctgagya atnaggaggt catygcctcc 900
ttggaggagg tcttgggaaa atacatgctg ccagcaactt ccacgctcta g 951

```

<210> 40
 <211> 1395
 <412> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472451CB1

```

<400> 40
atggaaggca catttgatcc tctaatggc attatgctt acsuaactct ggagagggat 60
ggcaaacgca aggaactgca aagtggaaac cagactctct tgcctcactt caatttgggg 120
ggccttccac acccaccaca gctgggagcg ccaactctct taacttctct tgtactctat 180
ctcctcactg tttctggaaa tgggtctcct atcctcactg tctttagtga catccggctc 240
cabctgcccc tctgcttgtt cctgtgtcac ctctcctctc tggacatgac caattcttgt 300
gctatgtccc ccaagatgct ggcctggctt ctcttgggta gtcaggatct ctctcttggg 360
ggctgtgaaa tcaactctat tcttctcact ctcttgggct gtcctgtagt ctctcttccc 420
accctatgag cttaagaccg ttctcttgcc atthtaagc ccttacctta lctaccatc 480
atgaccacaa gactctgtaa ctctctgctc ttggccactt gcttggaggt gactatccat 540
taactttccc aanaagtttt tgtatctcgg ctgacctctc gtggcccaa tgggtctgag 600
taactttctt gtcacatccc tgcctctctg ctctctgctc gctggcctac ggcctaccac 660
gagctgggca ccttctgaga cahtggcttc ctggcctcca cctgcttcah gctcactctc 720
actctctatg gctatattgt agctgacctc ctgccaatbc cctcagcaga tgggcgccc 780
agctctctct ccaactgtgc tgcctaccct acigtgtctc ttgttacta tglctctctc 840
acttctcttt accctgggac ttgttccaaq ggcctctctg atgggtgtgt agctgtcttt 900
tcaactgtca tcaactctct gcttaactcc actctctaca cactg-gcah caaagcaatg 960
aagcagactt tcaagagctt agggggccac aagcaagagc tagaagaact agagtgtgtt 1020
catacaactg tgaaggaca cagcttctcc acacagggac acaagagctc tggctatctt 1080
gggcaccatg acagtgaaga accacaagtc aatgaaggac agattgtgta gyaagtatga 1140
cttgggggtg tggaggtggg agtacaacct gatcaccagc aggtatgaga ggttccccag 1200
vacaacltaa agaaaaatta tatgacctc tcaagagatg gggaaaaagc actggacaat 1260
atccggcctc caataatana aactctcaat aactttagag caaaggggga ctctctctac 1320
ctgctgaagc argctctatga aaactctaca cctaaactaa gcaaatatcc tgaanaacta 1380
aatctttcc cctag

```

<210> 41
 <211> 972
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472456CB1

```

<400> 41
atgaaacctg aaaaatggac tcaagtaaca agctttgtcc ttctgggttt ccccagtagc 60
caactcaatc agtctctggt gttctctggy tcaatgggga cctacattgt aacagccaca 120
ggcaagctgc taatttatgt gctcagctgg atagacaaa gctctcaact acagatgta 180
ttcttctctc ggaattctc ctctctctgg ctgttctctg taactgtttt ggttccccag 240
atgctctctg tcaactctac ggggatctac acatctctat ttgtcagctg catcactcag 300
tcaactctct actctctctt aggcaccact gactctctcc tcttggctgt catctctctg 360
tatctctctc tggcaatctg ccgaccactc cgtatgaga cctctgatga tggcctctct 420
gttctccaac lagtctctgg ctctctgcta gctggatctc tctgggtctc ttgccccact 480
gtcctctctg ccagcctctc tttctgtctc cccaatggta ttgaccactt cttctgtgac 540

```

WO 01/66742

PCT/US01/06814

```

agtctgacct tgcctcagct ttcttctggg gacaccac: tgcctgaaact ggtggctttc 600
atgctctctc cgtttggtgt actgggctca ctgctctga cctcagtttc ctatgctcgc 660
attcttgcga ctgtttctcag ggcccttaca gctgctgagc gaaggaaagc gttttcact 720
tgcgctctgc atctttacagt ggtggctcgc atctatggca gttcactctt tctctacatt 780
cgtatctcag agctctcagtc caaactgctc aacaaaagty cctcctctct gagctgcctc 840
atcaacccc tcttgaccct atctcctctc actctctgca atgacaaggt gcagctagca 900
ctgagagag ccttgggggt gccacagctc actgctgtga tgaacctgag ggtcacaagt 960
caaaagaaat ga

```

```

<210> 42
<211> 957
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472457CB1

```

```

<400> 42
atgaatccag caaatcattc ccaggctgca ggatttgttc taactgggct ctctcaggtt 60
tggagcttc gtttggttt cttcactgtt ttctctctgt tctatttat gactgtatgt 120
ggaaaccttc ttatgttggc catagtgcac tccggaccac accctgacac aacctgttat 180
ttctctctgt gcaatctctc ttctctggac ttctgtctac ctctcctcac agcactctgg 240
atgctggttg acttctctc aggcacacct accattctct ttgctgctgt cctgactcaa 300
ctctctctct tccactctct tggggcctc aagatctctc tgcctgctgt catgctgtat 360
gacgctaca ttgcattctc ccagccctgt cactacagc ccattatgaa tctgactgtc 420
tctgactctc ttatggcagc ctcttgggtg gggggcttca tccactccat agcaagatt 480
gcattgacta tccagctgct attctgttgg cctgcaagc lggcaacct ttattgtgat 540
gtgctctcag tgaaccaatt ggcctgcaca gatacctctg tcttagact tctaatgttg 600
tttaacatg ccttgggtgc cctgctgtgt ttctgtgtgc ttctggtgta gtaacagca 660
ctgctagtca tgcctcgaag caactcagc gggggctgca gaaggctct gttactctgt 720
gcctctcaca ttctctgtgt gaacttaact ttctgtctct gcatctact ctatacaag 780
ctttctcaga cattcccatc gacacaagcc gtctctgtgc tatacaaat tctcaccctc 840
atgctgactc ctgctcctca taccctgagn aacaaaggag tgatctctgc catgagagag 900
ctgtgagga gaaaaaaggc ccttatgtgt cccctggagc acagacctt acattag 957

```

【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
13 September 2001 (13.09.2001)

PCT

(10) International Publication Number
WO 01/66742 A3

- (51) International Patent Classification: C12N 15/12, 15/10, 15/06, C07K 14/705, 16/28, A01K 67/027, A61K 38/17, C12Q 1/68, G01N 33/68
- (21) International Application Number: PCT/US01/09814
- (22) International Filing Date: 1 March 2001 (01.03.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/186,854 3 March 2000 (03.03.2000) US
60/188,384 10 March 2000 (10.03.2000) US
60/190,453 17 March 2000 (17.03.2000) US
60/190,730 20 March 2000 (20.03.2000) US
- (71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (72) Inventors; and
(73) Inventors/Applicants (for US only): LAI, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US); TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US); PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US); YAO, Monique, G. [US/US]; 111 Frederick Court, Mountain View, CA 94025 (US); SHIH, Leo, L. [US/US]; 1081 Tanford Drive, Apt. B, Palo Alto, CA 94303 (US); TRIBOULEY, Catherine, M. [FR/US]; 1121 Tennessee Street #5, San Francisco, CA 94107 (US); LU, Dyang, Anna, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US); YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US); KHAN, Farrah, A. [IN/US]; 3617 Central Road #102, Glenview, Illinois 60025 (US); POLICKY, Jennifer, J. [US/US]; 1511 Jarvis Court, San Jose, CA 95118 (US); AU-YOUNG, Janice [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US); YANG, Junming [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US); HARRIAND, Lee [GB/GB]; 18 Chaucer Court, New Dover Road, Canterbury Kent CT1 3AU (GB); WALSH, Roderick, T. [IE/GB]; 8 Boundary Court, St. Lawrence Road, Canterbury Kent CT1 3EZ (GB); LO, Terence, P. [CA/US]; 1453 Beach Park Boulevard, Apt. 115, Foster City, CA 94404-1941 (US); BOROVSKEY, Mark, L. [US/US]; 132 Orchard Avenue, Redwood City, CA 94061 (US).
- (74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (81) Designated States (national): AF, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, GR, HU, IL, ID, 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 (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).
- Published:
— with international search report
- (88) Date of publication of the international search report:
10 May 2002
- 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 01/66742 A3

(54) Title: G-PROTEIN COUPLED RECEPTORS

(57) Abstract: The invention provides human G-protein coupled receptors (GPCR) and polynucleotides which identify and encode GPCR. 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 GPCR.

【國際調查報告】

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/06814

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C12N15/10 C12N15/66 C07K14/705 C07K16/28
A01K67/027 A61K38/17 C12Q1/68 G01N33/68

According to International Patent Classification (IPC) or to be national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K A01K C12Q G01N

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)
EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99 57132 A (GENETICS INST) 11 November 1999 (1999-11-11) see seq. ID's 113 and 114, and clone mj301_1, and passages relating thereto.	1-19, 22, 25-28
X	WO 99 33869 A (CORIXA CORP) 8 July 1999 (1999-07-08) see seq. ID. 57	1, 3, 6-16, 18
A	WO 98 46620 A (MILLENNIUM PHARM INC) 22 October 1998 (1998-10-22) the whole document	
A	WO 99 63087 A (HODG MARTIN R ; GLUCKSMANN MARIA ALEXANDRA (US); MILLENNIUM PHARM I) 9 December 1999 (1999-12-09) the whole document	
	-/--	

Further documents are listed in the continuation of box C. 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

"I" document which may throw doubts on priority claims or which is cited to establish the publication date of another claim 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" large document published after the international filing date or priority date and not in conformance with the publication fee rules 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: 28 September 2001
Date of making of the international search report: 17 01 2002

Name and mailing address of the ISA: European Patent Office, P. B. 6818 Patentlaan 2, NL - 5280 HH Rijswijk, tel (+31-78) 940-2049, telex 31 654 4pp nld, fax (+31-78) 940-2049
Authorized officer: Smalt, R

2

Form PCT/ISA/210 (September 2001) (with 1993)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/06814

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Relevant to claim No.	
P,X	EP 1 074 617 A (HELIX RES INST) 7 February 2001 (2001-02-07) see seq.ID's 3981, 15315, and clone OVARC1001170. ---	3-8, 11-15,27
P,X	WO 00 61756 A (CORIXA CORP ; REED STEVEN G (US); DILLON DAVIN C (US); XU JIANGCHUN) 19 October 2000 (2000-10-19) See clone 1015F5, and seq.ID 57. -----	3,11-15

2

Form PCT/ISA 210 (continuation of sequence sheet) (July 1992)

INTERNATIONAL SEARCH REPORT		International application No. PCT/US 01/06814
Box I Observations where certain claims were found unsearchable (Continuation of item 3 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 type="checkbox"/>	Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
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:		
see additional sheet		
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.:
1-28, all 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 01/06814

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

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

1. Claims: Invention 1: claims 1-28, all partially

Isolated nucleic acid according to seq.ID.22, fragments or variants thereof, polypeptides encoded thereby, expression vector comprising said nucleic acid sequence, host or transgenic organism comprising said vector, method for producing said polypeptide using said host, antibody directed at said polypeptide, method for detecting said polynucleotide, composition comprising said polypeptide, method of screening for (antagonists and/or specifically binding compounds and/or activity modulators of said polypeptide, method of screening for compounds which alter expression of said nucleic acid, and use of said method to assess the toxicity of a compound.

2. Claims: Inventions 2-21: claims 1-28, all partially

Subject matter as defined for invention 1, but related to the respective G-protein coupled receptors with seq.ID's:

2. seq.ID's 2 and 23
3. seq.ID's 3 and 24
4. seq.ID's 4 and 25
5. seq.ID's 5 and 26
6. seq.ID's 6 and 27
7. seq.ID's 7 and 28
8. seq.ID's 8 and 29
9. seq.ID's 9 and 30
10. seq.ID's 10 and 31
11. seq.ID's 11 and 32
12. seq.ID's 12 and 33
13. seq.ID's 13 and 34
14. seq.ID's 14 and 35
15. seq.ID's 15 and 36
16. seq.ID's 16 and 37
17. seq.ID's 17 and 38
18. seq.ID's 18 and 39
19. seq.ID's 19 and 40
20. seq.ID's 20 and 41
21. seq.ID's 21 and 42.

International Application No. PCT/US 01/06814

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box 3.

Although claim 18 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Further defect(s) under Article 17(2)(a):

Continuation of Box 3.

Claims Nos.: 20,21,23,24

Present claims 20, 21, 22, and 24 relate to a compound defined by reference to a desirable characteristic or property, namely that it acts as an (ant)agonist for the G-protein(s) of claim 1. The claims cover all compounds having this characteristic or property, whereas the application provides no support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT for any such compounds. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search for said claims is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT). An attempt is made to define the compound by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search impossible. Consequently, said claims have not been searched.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.

PCT/US 01/06314

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9957132 A	11-11-1999	AU 4071199 A	23-11-1999
		EP 1077991 A1	28-02-2001
		WO 9957132 A1	11-11-1999
		AU 5475199 A	06-03-2000
		EP 1112285 A1	04-07-2001
		WO 0009551 A1	24-02-2000
WO 9933869 A	08-07-1999	AU 2010699 A	19-07-1999
		EP 1042360 A2	11-10-2000
		WO 9933869 A2	08-07-1999
		US 2001018058 A1	30-08-2001
		ZA 9811800 A	23-06-1999
WO 9846620 A	22-10-1998	US 5691720 A	06-04-1999
		AU 6973698 A	11-11-1998
		EP 1007536 A1	14-06-2000
		WO 9846620 A1	22-10-1998
WO 9963087 A	09-12-1999	AU 4544999 A	20-12-1999
		EP 1084241 A1	21-03-2001
		WO 9963087 A1	09-12-1999
EP 1074617 A	07-02-2001	AU 6180800 A	19-02-2001
		AU 6180900 A	19-02-2001
		AU 6181000 A	19-02-2001
		AU 6181100 A	19-02-2001
		AU 6181200 A	19-02-2001
		AU 6181300 A	19-02-2001
		AU 6181400 A	19-02-2001
		AU 6181500 A	19-02-2001
		AU 6181600 A	19-02-2001
		AU 6315800 A	19-02-2001
		EP 1074617 A2	07-02-2001
		WO 0109315 A1	08-02-2001
		WO 0109345 A1	08-02-2001
		WO 0109316 A1	08-02-2001
		WO 0109349 A1	08-02-2001
		WO 0109317 A1	08-02-2001
		WO 0109318 A1	08-02-2001
		WO 0109319 A1	08-02-2001
		WO 0109346 A1	08-02-2001
		WO 0109320 A1	08-02-2001
		WO 0109321 A1	08-02-2001
		WO 0109322 A1	08-02-2001
WO 0109323 A1	08-02-2001		
WO 0061756 A	19-10-2000	AU 4231300 A	14-11-2000
		EP 1169446 A2	09-01-2002
		WO 0061756 A2	19-10-2000
		US 2001018058 A1	30-08-2001

Form PCT/IB/210 (Patent family annex, July 1999)

フロントページの続き

(51) Int.Cl. ⁷	F I	テーマコード(参考)
A 6 1 P 1/04	A 6 1 P 1/08	4 C 0 8 4
A 6 1 P 1/08	A 6 1 P 1/10	4 H 0 4 5
A 6 1 P 1/10	A 6 1 P 1/12	
A 6 1 P 1/12	A 6 1 P 1/14	
A 6 1 P 1/14	A 6 1 P 1/16	
A 6 1 P 1/16	A 6 1 P 1/16	1 0 1
A 6 1 P 1/18	A 6 1 P 1/16	1 0 5
A 6 1 P 3/00	A 6 1 P 1/18	
A 6 1 P 3/04	A 6 1 P 3/00	
A 6 1 P 3/10	A 6 1 P 3/04	
A 6 1 P 7/00	A 6 1 P 3/10	
A 6 1 P 7/02	A 6 1 P 7/00	
A 6 1 P 7/06	A 6 1 P 7/02	
A 6 1 P 9/10	A 6 1 P 7/06	
A 6 1 P 9/12	A 6 1 P 9/10	
A 6 1 P 11/00	A 6 1 P 9/12	
A 6 1 P 11/06	A 6 1 P 11/00	
A 6 1 P 17/02	A 6 1 P 11/06	
A 6 1 P 17/06	A 6 1 P 17/02	
A 6 1 P 19/02	A 6 1 P 17/06	
A 6 1 P 19/04	A 6 1 P 19/02	
A 6 1 P 19/10	A 6 1 P 19/04	
A 6 1 P 21/02	A 6 1 P 19/10	
A 6 1 P 21/04	A 6 1 P 21/02	
A 6 1 P 25/00	A 6 1 P 21/04	
A 6 1 P 25/02	A 6 1 P 25/00	
A 6 1 P 25/04	A 6 1 P 25/02	
A 6 1 P 25/08	A 6 1 P 25/02	1 0 3
A 6 1 P 25/14	A 6 1 P 25/04	
A 6 1 P 25/16	A 6 1 P 25/08	
A 6 1 P 25/18	A 6 1 P 25/14	
A 6 1 P 25/20	A 6 1 P 25/16	
A 6 1 P 25/22	A 6 1 P 25/18	
A 6 1 P 25/28	A 6 1 P 25/20	
A 6 1 P 27/02	A 6 1 P 25/22	
A 6 1 P 29/00	A 6 1 P 25/28	
A 6 1 P 31/04	A 6 1 P 27/02	
A 6 1 P 31/10	A 6 1 P 29/00	
A 6 1 P 31/12	A 6 1 P 29/00	1 0 1
A 6 1 P 31/14	A 6 1 P 31/04	
A 6 1 P 31/16	A 6 1 P 31/10	
A 6 1 P 31/22	A 6 1 P 31/12	
A 6 1 P 33/00	A 6 1 P 31/14	
A 6 1 P 33/02	A 6 1 P 31/16	
A 6 1 P 35/00	A 6 1 P 31/22	
A 6 1 P 35/02	A 6 1 P 33/00	
A 6 1 P 37/02	A 6 1 P 33/02	

C 0 7 K	14/705	A 6 1 P	35/00	
C 0 7 K	16/28	A 6 1 P	35/02	
C 1 2 N	1/15	A 6 1 P	37/02	
C 1 2 N	1/19	C 0 7 K	14/705	
C 1 2 N	1/21	C 0 7 K	16/28	
C 1 2 N	5/10	C 1 2 N	1/15	
C 1 2 P	21/02	C 1 2 N	1/19	
C 1 2 Q	1/68	C 1 2 N	1/21	
G 0 1 N	33/15	C 1 2 P	21/02	C
G 0 1 N	33/50	C 1 2 Q	1/68	A
G 0 1 N	33/53	C 1 2 Q	1/68	Z
G 0 1 N	33/566	G 0 1 N	33/15	Z
		G 0 1 N	33/50	Z
		G 0 1 N	33/53	M
		G 0 1 N	33/566	
		C 1 2 N	5/00	A
		A 6 1 K	37/02	

(31)優先権主張番号 60/190,730

(32)優先日 平成12年3月20日(2000.3.20)

(33)優先権主張国 米国(US)

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,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,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,S,G,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZW

(72)発明者 タング、ワイ・トム

アメリカ合衆国カリフォルニア州9 5 1 1 8・サンノゼ・ランウィックコート 4 2 3 0

(72)発明者 パターソン、チャンドラ

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・シャーウッドウェイ 4 9 0

(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 2 5・マウンテンビュー・フレデリックコート 1 1 1

(72)発明者 シー、レオ・エル

アメリカ合衆国カリフォルニア州9 4 3 0 3・パロアルト・アパートメント ビー・タンランドドライブ 1 0 8 1

(72)発明者 トリボレー、キャサリン・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 リュ、デュング・アイナ・エム

アメリカ合衆国カリフォルニア州9 5 1 2 3・サンノゼ・コイドライブ 2 3 3

(72)発明者 ユエ、ヘンリー

アメリカ合衆国カリフォルニア州9 4 0 8 7・サニーベイル・ルイスアベニュー 8 2 6

(72)発明者 カーン、ファラ・エイ

アメリカ合衆国イリノイ州6 0 0 2 5・グレンビュー・# 1 0 2・セントラルロード 3 6 1 7

(72)発明者 ポリッキー、ジェニファー・エル

アメリカ合衆国カリフォルニア州9 5 1 1 8・サンノゼ・ジャービスコート 1 5 1 1

(72)発明者 オウ・ヤング、ジャニス

アメリカ合衆国カリフォルニア州9 4 0 0 5・プリズペーン・ゴールドデンイーグルレーン 2 3 3

- (72)発明者 ヤング、ジュンミング
アメリカ合衆国カリフォルニア州 9 5 1 2 9 ・サンノゼ・パークレーン 7 1 2 5
- (72)発明者 ハーランド、リー
イギリス国ケント州・シーティー 1 3 エイユー・カンタベリー・ニュードーバーロード・チョーサーコート 1 8
- (72)発明者 ウォルシュ、ロドリック・ティー
イギリス国ケント州・シーティー 1 3 イージー・カンタベリー・セントローレンスロード・バウンダリーコート 8
- (72)発明者 ロー、テレンス・ピー
アメリカ合衆国カリフォルニア州 9 4 4 0 4 - 1 9 4 1 ・フォスターシティ・アパートメント 1 1 5 ・ビーチパークプールバード 1 4 5 2
- (72)発明者 ボロースキー、マーク・エル
アメリカ合衆国カリフォルニア州 9 4 0 6 1 ・レッドウッドシティ・オーチャードアベニュー 1 2 2

F ターム(参考) 2G045 AA25 AA40 BA13 BB20 CB01 CB21 DA12 DA13 DA14 DA36
DA77 FB02 FB03 FB07

4B024 AA01 AA11 BA61 BA63 CA04 CA11 DA01 DA02 DA05 DA11
EA02 EA04 GA01 GA11 GA18 GA19 HA08 HA11 HA12

4B063 QA01 QA18 QQ02 QQ08 QQ42 QQ52 QR08 QR14 QR20 QR31
QR42 QR50 QR56 QR58 QR62 QR72 QX02

4B064 AG20 AG26 CA02 CA05 CA10 CA11 CA12 CA19 CA20 CC24
DA03 DA13

4B065 AA01X AA57X AA87X AA93Y AB01 BA01 BA08 CA23 CA24 CA44
CA46

4C084 AA01 AA16 NA14 ZA01 ZA02 ZA05 ZA06 ZA12 ZA15 ZA16
ZA18 ZA20 ZA22 ZA24 ZA33 ZA36 ZA37 ZA38 ZA39 ZA40
ZA42 ZA51 ZA53 ZA54 ZA59 ZA60 ZA66 ZA68 ZA69 ZA70
ZA71 ZA72 ZA73 ZA75 ZA77 ZA89 ZA94 ZB07 ZB08 ZB09
ZB15 ZB26 ZB27 ZB33 ZB37 ZB38 ZB39 ZC21 ZC35

4H045 AA10 AA11 AA20 AA30 BA09 CA40 DA50 DA75 EA20 EA50
FA72 FA73 FA74

专利名称(译)	G蛋白偶联受体		
公开(公告)号	JP2004507212A	公开(公告)日	2004-03-11
申请号	JP2001565896	申请日	2001-03-01
[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	ラルプリーティ タングワイトム パターソンチャンドラ ヤオモニークジー シーレオエル トリボレーキャサリーンエム リュデュングアイナエム ユエハンリー カーンファラエイ ポリッキージェニファーエル オウヤングジャンス ヤングジュンミング ハーランドリー ウォルシュロドリックティー ローテレンスピー ボロースキーマークエル		
发明人	ラル、プリーティ タング、ワイトム パターソン、チャンドラ ヤオ、モニークジー シー、レオエル トリボレー、キャサリーン・エム リュ、デュング・アイナ・エム ユエ、ハンリー カーン、ファラ・エイ ポリッキー、ジェニファー・エル オウ・ヤング、ジャンス ヤング、ジュンミング ハーランド、リー ウォルシュ、ロドリック・ティー ロー、テレンスピー ボロースキー、マーク・エル		
IPC分类号	A01K67/027 A61K38/00 A61K45/00 A61P1/00 A61P1/04 A61P1/08 A61P1/10 A61P1/12 A61P1/14 A61P1/16 A61P1/18 A61P3/00 A61P3/04 A61P3/10 A61P7/00 A61P7/02 A61P7/06 A61P9/10 A61P9 /12 A61P11/00 A61P11/06 A61P17/02 A61P17/06 A61P19/02 A61P19/04 A61P19/10 A61P21/02 A61P21/04 A61P25/00 A61P25/02 A61P25/04 A61P25/08 A61P25/14 A61P25/16 A61P25/18 A61P25 /20 A61P25/22 A61P25/28 A61P27/02 A61P29/00 A61P31/04 A61P31/10 A61P31/12 A61P31/14 A61P31/16 A61P31/22 A61P33/00 A61P33/02 A61P35/00 A61P35/02 A61P37/02 C07K14/705 C07K16 /28 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N15/09 C12N15/12 C12P21/02 C12Q1/68 G01N33 /15 G01N33/50 G01N33/53 G01N33/566		
CPC分类号	A01K2217/05 A61K38/00 A61P1/00 A61P1/04 A61P1/08 A61P1/10 A61P1/12 A61P1/14 A61P1/16 A61P1/18 A61P3/00 A61P3/04 A61P3/10 A61P7/00 A61P7/02 A61P7/06 A61P9/10 A61P9/12 A61P11		

/00 A61P11/06 A61P17/02 A61P17/06 A61P19/02 A61P19/04 A61P19/10 A61P21/02 A61P21/04
 A61P25/00 A61P25/02 A61P25/04 A61P25/08 A61P25/14 A61P25/16 A61P25/18 A61P25/20 A61P25
 /22 A61P25/28 A61P27/02 A61P29/00 A61P31/04 A61P31/10 A61P31/12 A61P31/14 A61P31/16
 A61P31/22 A61P33/00 A61P33/02 A61P35/00 A61P35/02 A61P37/02 C07K14/705

FI分类号 C12N15/00.ZNA.A A01K67/027 A61K45/00 A61P1/00 A61P1/04 A61P1/08 A61P1/10 A61P1/12 A61P1
 /14 A61P1/16 A61P1/16.101 A61P1/16.105 A61P1/18 A61P3/00 A61P3/04 A61P3/10 A61P7/00 A61P7
 /02 A61P7/06 A61P9/10 A61P9/12 A61P11/00 A61P11/06 A61P17/02 A61P17/06 A61P19/02 A61P19
 /04 A61P19/10 A61P21/02 A61P21/04 A61P25/00 A61P25/02 A61P25/02.103 A61P25/04 A61P25/08
 A61P25/14 A61P25/16 A61P25/18 A61P25/20 A61P25/22 A61P25/28 A61P27/02 A61P29/00 A61P29
 /00.101 A61P31/04 A61P31/10 A61P31/12 A61P31/14 A61P31/16 A61P31/22 A61P33/00 A61P33/02
 A61P35/00 A61P35/02 A61P37/02 C07K14/705 C07K16/28 C12N1/15 C12N1/19 C12N1/21 C12P21
 /02.C C12Q1/68.A C12Q1/68.Z G01N33/15.Z G01N33/50.Z G01N33/53.M G01N33/566 C12N5/00.A
 A61K37/02

F-TERM分类号 2G045/AA25 2G045/AA40 2G045/BA13 2G045/BB20 2G045/CB01 2G045/CB21 2G045/DA12 2G045
 /DA13 2G045/DA14 2G045/DA36 2G045/DA77 2G045/FB02 2G045/FB03 2G045/FB07 4B024/AA01
 4B024/AA11 4B024/BA61 4B024/BA63 4B024/CA04 4B024/CA11 4B024/DA01 4B024/DA02 4B024
 /DA05 4B024/DA11 4B024/EA02 4B024/EA04 4B024/GA01 4B024/GA11 4B024/GA18 4B024/GA19
 4B024/HA08 4B024/HA11 4B024/HA12 4B063/QA01 4B063/QA18 4B063/QQ02 4B063/QQ08 4B063
 /QQ42 4B063/QQ52 4B063/QR08 4B063/QR14 4B063/QR20 4B063/QR31 4B063/QR42 4B063/QR50
 4B063/QR56 4B063/QR58 4B063/QR62 4B063/QR72 4B063/QX02 4B064/AG20 4B064/AG26 4B064
 /CA02 4B064/CA05 4B064/CA10 4B064/CA11 4B064/CA12 4B064/CA19 4B064/CA20 4B064/CC24
 4B064/DA03 4B064/DA13 4B065/AA01X 4B065/AA57X 4B065/AA87X 4B065/AA93Y 4B065/AB01
 4B065/BA01 4B065/BA08 4B065/CA23 4B065/CA24 4B065/CA44 4B065/CA46 4C084/AA01 4C084
 /AA16 4C084/NA14 4C084/ZA01 4C084/ZA02 4C084/ZA05 4C084/ZA06 4C084/ZA12 4C084/ZA15
 4C084/ZA16 4C084/ZA18 4C084/ZA20 4C084/ZA22 4C084/ZA24 4C084/ZA33 4C084/ZA36 4C084
 /ZA37 4C084/ZA38 4C084/ZA39 4C084/ZA40 4C084/ZA42 4C084/ZA51 4C084/ZA53 4C084/ZA54
 4C084/ZA59 4C084/ZA60 4C084/ZA66 4C084/ZA68 4C084/ZA69 4C084/ZA70 4C084/ZA71 4C084
 /ZA72 4C084/ZA73 4C084/ZA75 4C084/ZA77 4C084/ZA89 4C084/ZA94 4C084/ZB07 4C084/ZB08
 4C084/ZB09 4C084/ZB15 4C084/ZB26 4C084/ZB27 4C084/ZB33 4C084/ZB37 4C084/ZB38 4C084
 /ZB39 4C084/ZC21 4C084/ZC35 4H045/AA10 4H045/AA11 4H045/AA20 4H045/AA30 4H045/BA09
 4H045/CA40 4H045/DA50 4H045/DA75 4H045/EA20 4H045/EA50 4H045/FA72 4H045/FA73 4H045
 /FA74

優先権 60/186854 2000-03-03 US
 60/188384 2000-03-10 US
 60/190453 2000-03-17 US
 60/190730 2000-03-20 US

外部リンク Espacenet

摘要(译)

本发明提供了鉴定和编码GCREC的人G蛋白偶联受体 (GCREC) 和多核苷酸。 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GCREC异常表达有关的疾病的方法。

		(P2004-507212A) (43) 公表日 平成16年3月11日(2004.3.11)	
(51) Int. Cl. ⁷	F I	テーマコード (参考)	
C 1 2 N 15/09	C 1 2 N 15/00	Z N A A	2 G 0 4 5
A 0 1 K 67/027	A 0 1 K 67/027		4 B 0 2 4
A 6 1 K 38/00	A 6 1 K 45/00		4 B 0 6 3
A 6 1 K 45/00	A 6 1 P 1/00		4 B 0 6 4
A 6 1 P 1/00	A 6 1 P 1/04		4 B 0 6 5
	審査請求 未請求 予備審査請求 有 (全 251 頁) 最終頁に続く		
(21) 出願番号	特願2001-565896 (P2001-565896)	(71) 出願人	301005050
(60) (22) 出願日	平成13年3月1日 (2001.3.1)		インサイト・ゲノミクス・インコーポレ イテッド
(83) 翻訳文提出日	平成14年9月2日 (2002.9.2)		アメリカ合衆国カリフォルニア州9430 4・パロアルト・ボートドライブ 31 60
(86) 国際出願番号	PCT/US2001/006814	(74) 代理人	100089266
(87) 国際公開番号	W02001/066742		弁理士 大島 陽一 ラル、フリーティ
(87) 国際公開日	平成13年9月13日 (2001.9.13)	(72) 発明者	アメリカ合衆国カリフォルニア州9505 4・サンタクララ・ラスドライブ 238 2
(31) 優先権主張番号	60/186,854		
(32) 優先日	平成12年3月3日 (2000.3.3)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/188,384		
(32) 優先日	平成12年3月10日 (2000.3.10)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/190,453		
(32) 優先日	平成12年3月17日 (2000.3.17)		
(33) 優先権主張国	米国 (US)		

最終頁に続く