

(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2004-500808
(P2004-500808A)

(43) 公表日 平成16年1月15日(2004.1.15)

(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 31/711	A 6 1 K 31/711	4 B O 6 3
A 6 1 K 38/00	A 6 1 K 45/00	4 B O 6 4
A 6 1 K 45/00	A 6 1 K 48/00	4 B O 6 5
	審査請求 未請求 予備審査請求 有 (全 295 頁) 最終頁に続く	

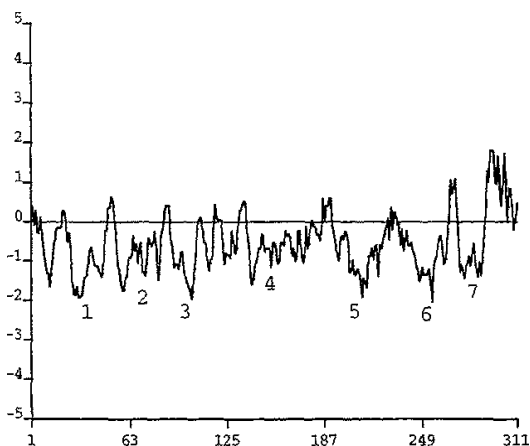
(21) 出願番号	特願2001-543585 (P2001-543585)	(71) 出願人	301005050 インサイト・ゲノミックス・インコーポレイテッド アメリカ合衆国カリフォルニア州94304・パロアルト・ポータードライブ 3160
(86) (22) 出願日	平成12年12月7日 (2000.12.7)	(74) 代理人	100089266 弁理士 大島 陽一
(85) 翻訳文提出日	平成14年6月10日 (2002.6.10)	(72) 発明者	バーフォード、ニール アメリカ合衆国コネチカット州06422・ダラム・ワイルドウッドサークル 105
(86) 国際出願番号	PCT/US2000/033382		
(87) 国際公開番号	W02001/042288		
(87) 国際公開日	平成13年6月14日 (2001.6.14)		
(31) 優先権主張番号	60/172, 852		
(32) 優先日	平成11年12月10日 (1999.12.10)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/171, 732		
(32) 優先日	平成11年12月22日 (1999.12.22)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/176, 148		
(32) 優先日	平成12年1月14日 (2000.1.14)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 Gタンパク質結合受容体

(57) 【要約】

本発明は、ヒトGタンパク質結合受容体 (GCREC) と、GCRECを同定及びコードするポリヌクレオチドを提供する。本発明はまた、発現ベクター及び宿主細胞、抗体、アゴニスト、アンタゴニストを提供する。更に、本発明は、GCRECの発現に関連する疾患の診断・治療・予防方法を提供する。



SEQ ID NO:1
(Incyte ID No. 104941CD1)

【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: 39 (SEQ ID NO: 1 - 39) からなる群から選択されたアミノ酸配列と、

(b) SEQ ID NO: 1 - 39 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 - 39 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 - 39 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 - 39 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: 40 - 78 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: 40 - 78 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: 40 - 78 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 1 3】

サンプルにおいて、請求項 1 1 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 2 0 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 1 4】

前記プローブが少なくとも 6 0 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 1 3 に記載の方法。

【請求項 1 5】

サンプルにおいて、請求項 1 1 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、 20

(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。

【請求項 1 6】

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 1 7】

前記ポリペプチドが、SEQ ID NO : 1 - 3 9 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 1 6 の組成物。

【請求項 1 8】

機能的な G C R E C (新規の G タンパク質結合受容体) 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 1 6 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 1 9】

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 2 0】

請求項 1 9 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 2 1】

機能的な G C R 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 2 0 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 2 2】

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 2 3】

請求項 2 2 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 2 4】

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 2 3 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 2 5】

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 2 6】

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 2 7】

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 2 8】

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 1 1 のポリヌクレオチドの少なくとも 2 0 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 1 1 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タンパク質結合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タンパク質結合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効果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タンパク質結合受容体は、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

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関与する(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関与する。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タンパク質結合受容体、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タンパク質結合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効果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」、「GCREC-21」、「GCREC-22」、「GCREC-23」、「GCREC-24」、「GCREC-25」、「GCREC-26」、「GCREC-27」、「GCREC-28」、「GCREC-29」、「GCREC-30」、「GCREC-31」、「GCREC-32」、「GCREC-33」、「GCREC-34」、「GCREC-35」、「GCREC-36」、「GCREC-37」、「GCREC-38」、および「GCREC-39」と呼ぶGタンパク質結合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乃至21 (SEQ ID NO: 1-39) からなる一群から選択されたアミノ酸配列と、(b)SEQ ID NO: 1-39 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: 1-39 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: 1-39 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: 1-39 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: 1-39 からなる一群から選択されたアミノ酸配列と、(b)SEQ ID NO: 1-39 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: 1-39 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: 1-39 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: 1-39 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: 40-78 からなる一群から選択される。

【0020】

更に、本発明は、(a)SEQ ID NO: 1-39 からなる一群から選択されたアミノ酸配列と、(b)SEQ ID NO: 1-39 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: 1-39 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: 1-39 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。

10

20

30

40

50

【0021】

また、本発明は、(a) SEQ ID NO: 1 - 39 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 - 39 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 - 39 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 - 39 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。

【0022】

更に、本発明は、(a) SEQ ID NO: 1 - 39 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 - 39 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 - 39 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 - 39 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。

【0023】

更に、本発明は、(a) SEQ ID NO: 40 - 78 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) SEQ ID NO: 40 - 78 からなる一群から選択されたポリヌクレオチド配列と90%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RNA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60個の連続するヌクレオチドを含む。

【0024】

更に本発明は、(a) SEQ ID NO: 40 - 78 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) SEQ ID NO: 40 - 78 からなる一群から選択されたポリヌクレオチド配列と90%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RNA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20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60個の連続するヌクレオチドを含む。

【0025】

更に本発明は、(a) SEQ ID NO: 40 - 78 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) SEQ ID NO: 40 - 78 からなる一群から選択されたポリヌクレオチド配列と90%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RNA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。

【0026】

更に本発明は、(a)SEQ ID NO: 1-39からなる一群から選択されたアミノ酸配列と、(b)SEQ ID NO: 1-39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: 1-39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: 1-39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: 1-39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。

10

【0027】

更に本発明は、(a)SEQ ID NO: 1-39からなる一群から選択されたアミノ酸配列と、(b)SEQ ID NO: 1-39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: 1-39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: 1-39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。

20

【0028】

更に、本発明は、(a)SEQ ID NO: 1-39からなる一群から選択されたアミノ酸配列と、(b)SEQ ID NO: 1-39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: 1-39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: 1-39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

40

【0029】

更に本発明は、(a)SEQ ID NO: 1-39からなる一群から選択されたアミノ酸配列と、(b)SEQ ID NO: 1-39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: 1-39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: 1-39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

50

ステップとを含む。

【0030】

更に本発明は、(a) SEQ ID NO: 1 - 39 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 - 39 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 - 39 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 - 39 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: 40 - 78 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。

20

【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: 40 - 78 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: 40 - 78 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: 40 - 78 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: 40 - 78 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。

30

40

【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

50

たものではないということも理解されたい。

【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。従来の特許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。

10

【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たり、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。

20

【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。

30

【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適當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。

40

【0040】

50

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、アミノ酸配列を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。

【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、GCREC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、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。GCREC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（benzylphosphonate）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GCREC、合成のGCREC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。G C R E C 若しくは G C R E C 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】

10

「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。

【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。

20

元の残基

保存的な置換

Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
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

30

40

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。

【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】

50

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。

10

【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（parent sequence）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。

20

【0055】

SEQ ID NO: 40 - 78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、SEQ ID NO: 40 - 78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。SEQ ID NO: 40 - 78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SEQ ID NO: 40 - 78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SEQ ID NO: 40 - 78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。

30

【0056】

SEQ ID NO: 1 - 39のある断片は、SEQ ID NO: 40 - 78のある断片によってコードされる。SEQ ID NO: 1 - 39のある断片は、特異的にSEQ ID NO: 1 - 39を同定する固有のアミノ酸配列の領域を含む。例えば、SEQ ID NO: 1 - 39のある断片は、免疫原性ペプチドとして有用である。

【0057】

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。

40

【0058】

ポリヌクレオチド配列間の同一性のパーセントは、MEGALIGN version 3.12e配列アラインメントプログラムに組込まれる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。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

50

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、`K t u p l e = 2`、`g a p p e n a l t y = 5`、`w i n d o w = 4`、「`d i a g o n a l s s a v e d`」= 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。

【0059】

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。

【0060】

```
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。

【0061】

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。

【0062】

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。

【0063】

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ンメントの場合、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

10

【0064】

別法では、NCBI BLASTソフトウェア式が用いられる。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。

【0065】

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

20

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

【0066】

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

【0067】

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。

【0068】

「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100µg/m

40

50

1のせん断して変性したサケ精子DNAが含まれる。

【0069】

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

【0070】

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0.2x 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0.1~2x 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 $\mu\text{g}/\text{ml}$ 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

【0071】

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、フィルター、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得る。

30

【0072】

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。

【0073】

「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。

【0074】

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GVRED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GCRECのポリペプチド断片またはオリゴペプチド断片を含む。

40

【0075】

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。

【0076】

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。

【0077】

用語「調節」は、GCRECの活性の変化を指す。例えば、調節によって、GCRECの

50

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。

【0078】

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。

【0079】

「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。

10

【0080】

「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。

【0081】

GCREC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、GCRECの酵素環境に依存し、細胞の種類によって異なり得る。

20

【0082】

「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、GCREC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

【0083】

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

【0084】

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 Laboratory 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、199

50

0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。

【0085】

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。

【0086】

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。

【0087】

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。

【0088】

「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

10

20

30

40

50

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。

【0089】

「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。

【0090】

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。

【0091】

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。

【0092】

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。

【0093】

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。

【0094】

「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。

【0095】

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、膜及びフィルター、チップ、スライド、ウェハ、ファイバー、磁気または非磁気ビード、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。

【0096】

「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。

【0097】

「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

10

20

30

40

50

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。

【0098】

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(*trans conjugation*)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。

10

【0099】

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%、95%、98%、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。

20

30

【0100】

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、ある核酸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%、85%、90%、95%、98%、或いはそれ以上の同一性を示し得る。

40

【0101】

(発明)

本発明は、新規のヒトGタンパク質結合受容体(GCRC)及びGCRCをコードするポリヌクレオチドの発見に基づき、これらの組成物を利用した増殖異常、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。

【0102】

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

50

番号 (Incyte Project ID) に相関する。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。

【0103】

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。

10

【0104】

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。

20

【0105】

表4に示されているように、本発明の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 (Incyte 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 : 40 - 78を同定するため、或いはSEQ ID NO : 40 - 78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。

30

40

【0106】

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、927003T6はインサイトcDNA配列の識別番号であり、BRAINOT04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、70489898V1) 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

50

cDNAすなわちEST(例えば、g835247)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、g4190944 v113. gs_10. editは、Genscan推定コード配列の識別番号であって、g4190944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(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。

10

【0107】

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。

【0108】

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GCREC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。

20

【0109】

本発明はまた、GCREC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、GCRECをコードするSEQ ID NO: 40-78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SEQ ID NO: 40-78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RNA配列を含む。

30

【0110】

本発明はまた、GCREC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、GCREC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、SEQ ID NO: 40-78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SEQ ID NO: 40-78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、GCREC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。

40

【0111】

遺伝暗号の縮重により作り出され得るGCREC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GCREC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。

50

【0112】

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。

10

【0113】

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。

【0114】

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、SEQ ID NO: 40 - 78 及びそれらの断片とハイブリダイズ可能なポリヌクレオチド配列が含まれる（例えば、Wahl, G. M. 及び S. L. Berger (1987) *Methods Enzymol.* 152: 399 - 407; and 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。

20

【0115】

当分野で周知の D N A 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 D N A ポリメラーゼ I のクレノウ断片、S E Q U E N A S E (U S B i o c h e m i c a l , C l e v e l a n d O H)、T a q ポリメラーゼ (A p p l i e d B i o s y s t e m s)、熱安定性 T 7 ポリメラーゼ (A m e r s h a m , P h a r m a c i a B i o t e c h P i s c a t a w a y N J)、或いは E L O N G A S E 増幅システム (L i f e T e c h n o l o g i e s , G a i t h e r s b u r g M D) 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、M I C R O L A B 2 2 0 0 液体転移システム (H a m i l t o n , R e n o , N V)、P T C 2 0 0 T h e r m a l C y c l e r 2 0 0 (M J R e s e a r c h , W a t e r t o w n M A) 及び A B I C A T A L Y S T 8 0 0 (P E B i o s y s t e m s) などの装置を用いて配列の準備を自動化する。次に、A B I 3 7 3 或いは 3 7 7 D N A シーケンシングシステム (P E B i o s y s t e m s)、M E G A B A C E 1 0 0 0 D N A シーケンシングシステム (M o l e c u l a r D y n a m i c s , S u n n y v a l e C A) 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）。

30

40

【0116】

当分野で周知の P C R 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、G C R E C 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 P C R 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 (n e s t e d p r i m e r) を用いてクローニン

50

ゲ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.ら(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 Bioscience, 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。

10

20

【0117】

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。

【0118】

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。

30

【0119】

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のクローン化及び発現に利用可能である。

40

【0120】

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ン化、プロセッシング及び/または発現の調節が含まれるが、これらに限定されるものではない。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

50

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、スプライスバリエーションの作製等が可能である。

【0121】

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、同じ或いは異なった種と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

【0122】

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.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。

30

【0123】

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)。

40

【0124】

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

50

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)。

10

【0125】

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)。

20

【0126】

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) *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、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(Dinicola, 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。

30

40

50

【0127】

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。

10

【0128】

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 12: 181 - 184. を参照)。

20

【0129】

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る。これらの作製物は、直接の DNA 形質転換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)。

30

【0130】

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。

40

【0131】

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

50

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を作製し、従来の輸送方法(リボソーム、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)。

【0132】

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。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。

10

【0133】

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はクロルスルホン(cGCRECsulfuron)、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)。アニトシアニン、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)。

20

30

【0134】

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。

40

【0135】

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

50

- 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。

【0136】

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 GC REC 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。GC REC 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)。

【0137】

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。GC REC 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、GC REC 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。

【0138】

GC REC 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。GC REC 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 GC REC 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。

【0139】

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。

【0140】

本発明の別の実施例では、GCREC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GCRECタンパク質が、GCREC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、GCREC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、GCREC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ch10）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。

10

20

【0141】

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。

30

【0142】

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。

【0143】

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。

40

【0144】

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

50

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。

【0145】

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。

【0146】

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、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「ロックアウト」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ロックアウトする(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。

【0147】

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)。

【0148】

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「ロックイン」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。ロ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)。

【0149】

(治療)

GCRECのある領域とGタンパク質結合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。更に、GCRECの発現は、癌組織、神経組織、胃腸組織、および肺組織に密接に関連する。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。

10

【0150】

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

20

30

40

50

、洞または幽門の浮腫、腹部アンギナ、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（A I D S）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（A I D S）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（A P E C E D）、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、糖尿病、肺気腫、リンパ球毒素性一時性リンパ球減少症、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、瘰癧、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、プンヤウイルス、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。

10

20

【0151】

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。

30

【0152】

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。

【0153】

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。

【0154】

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。

40

【0155】

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

50

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。

【0156】

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て投与することもできる。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。

【0157】

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。

【0158】

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、プルロニックポリオール、ポリアニオン、ペプチド、油性乳剤、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、B C G (b a c i l l i C a l m e t t e - G u e r i n) 及び C o r y n e b a c t e r i u m p a r v u m が特に好ましい。

【0159】

G C R E C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。G C R E Cアミノ酸の短いストレッチは、K L Hなどの別のタンパク質の配列と融合し、キメラ分子に対する抗体が産生され得る。

【0160】

G C R E C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E B V - 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、K o h l e r , G . ら . (1 9 7 5) N a t u r e 2 5 6 : 4 9 5 - 4 9 7 ; K o z b o r , D . ら . (1 9 8 5) J . I m m u n o l . M e t h o d s 8 1 - 8 - 4 2 ; C o t e , R . J . ら . (1 9 8 3) P r o c . N a t l . A c a d . S c i . 8 0 : 2 0 2 6 - 2 0 3 0 ; C o l e , S . P . ら . (1 9 8 4) M o l . C e l l B i o l . 6 2 : 1 0 9 - 1 2 0 を参照）。

【0161】

更に、「キメラ抗体」作製のために発達したヒト抗体遺伝子にマウス抗体遺伝子をスプライシングするなどの技術が、好適な抗原特異性及び生物学的活性を備える分子を得るために用いられる（例えば、M o r r i s o n , S . L . 他 . (1 9 8 4) P r o c . N a t l . A c a d . S c i . 8 1 - 4 8 5 1 - 4 8 5 5 ; N e u b e r g e r , M . S . 他 . (1 9 8 4) N a t u r e 3 1 2 : 6 0 4 - 6 0 8 ; T a k e d a , 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ィ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)。

【0162】

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)。

10

【0163】

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. 他。(1989) Science 254:1275-1281を参照)。

20

【0164】

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、GCREC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GCREC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(Pound、前出)。

【0165】

ラジオイムノアッセイ技術と共にScatchard分析などの様々な方法を用いて、GCRECに対する抗体の親和性を評価する。親和性を結合定数Kaで表すが、このKaは、平衡状態の下でGCREC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GCREC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Kaは、GCREC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GCRECエピトープに単一特異的なモノクローナル抗体医薬のKaは、親和性の真の測定値を表す。Ka値が $10^9 \sim 10^{12}$ L/molの高親和性抗体医薬は、GCREC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。Ka値が $10^6 \sim 10^7$ L/molの低親和性抗体医薬は、GCREC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。(Catty, D. (1988) Antibodies, Volume I: A Practical Approach. IRL Press, Washington, DC; Liddell, J. E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY)。

30

40

【0166】

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1~2 mg/mlの特異的な抗体、好ましくは5~10 mg/mlの特異的な抗体を含むポリクローナル

50

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 l i g a n 他、前出を参照)。

【0167】

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。

10

【0168】

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 . A l l e r g y C l i n . I m m u n o l . 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) B l o o d 76:271; A u s u b e l , 前出; U c k e r t , W . and W . W a l t h e r (1994) P h a r m a c o l . T h e r . 63(3):323-347を参照)。その他の遺伝子送達機構には、リポソーム系、人工的なウイルスエンベロープ、及び当分野で周知のその他の系が含まれる(R o s s i , J . J . (1995) B r . M e d . B u l l . 51(1):217-225; B o a d o , R . J . 他(1998) J . P h a r m . S c i . 87(11):1308-1315; and M o r r i s , M . C . 他(1997) N u c l e i c A c i d s R e s . 25(14):2730-2736.を参照)。

20

【0169】

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(C a v a z z a n a - C a l v o , M . 他(2000) S c i e n c e 288:669-672)によって特徴づけられる重度の複合型免疫欠損(S C I D) - X1)、遺伝性アデノシン-デアミナーゼ(A D A)欠損症(B l a e s e , R . M . 他(1995) S c i e n c e 270:475-480; B o r d i g n o n , C . 他(1995) S c i e n c e 270:470-475)に関連する重度の複合型免疫欠損、嚢胞性繊維症(Z a b n e r , J . 他(1993) C e l l 75:207-216; C r y s t a l , R . G . 他(1995) H u m . G e n e T h e r a p y 6:643-666; C r y s t a l , R . G . 他(1995) H u m . G e n e T h e r a p y 6:667-703)、サラセミア(t h a l a s s a m i a)、家族性高コレステロール血症、第V I I I因子若しくは第I X因子欠損による血友病(C r y s t a l , R . G . (1995) S c i e n c e 270:404-410; V e r m a , I . M . and S o m i a . N . (1997) N a t u r e 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(H I V)(B a l t i m o r e , D . (1988) N a t u r e 335:395-396; P o e s c h l a , E . 他(1996) P r o c . N a t l . A c a d . S c i . U S A . 93:11395-11399)や、B型若しくはC型肝炎ウイルス(H B V、H C V)、C a n d i

30

40

50

da 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。

【0170】

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トランスポゾン(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。

【0171】

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、PCDNA3.1、EPITAG、PRCCMV2、PREP、PVAXベクター(Invitrogen, Carlsbad CA)、PCMV-SCRIPT、PCMV-TAG、PEGSH/PERV(Stratagene, La Jolla CA)、PTET-OFF、PTET-ON、PTRE2、PTRE2-LUC、PTK-HYG(Clontech, Palo Alto CA)が含まれる。GCRECを発現させるために、(i)恒常的に活性化プロモーター(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。

【0172】

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。

【0173】

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

10

20

30

40

50

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GCRC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(Ranga, 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; Ranga, U. 他(1998) Proc. Natl. Acad. Sci. U.S.A. 95:1201-1206; Su, L. (1997) Blood 89:2283-2290)。

【0174】

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、GCRC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GCRC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隣臓の無損傷の膵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。

【0175】

別法では、ヘルペス系遺伝子治療の送達系を用いて、GCRC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GCRC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GCRC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。

10

【0176】

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) 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)。

20

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。

30

【0177】

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 Immunological GCREC 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。

40

50

【0178】

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。

【0179】

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。

10

【0180】

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、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。

20

【0181】

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。

30

【0182】

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関連する疾患の治療においては、GCRECをコードするポリヌクレオチドの発現を特異的に阻害する化合物が治療上有用であり、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をコードするポリヌクレオチドの発現を特異的に促進する化合物が治療上有用であり得る。

40

【0183】

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

50

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/または構造的な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)。

【0184】

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)。

【0185】

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。

【0186】

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。

【0187】

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。

【0188】

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、針注射を用いなくて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。

【0189】

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。

10

【0190】

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ポソーム製剤は、細胞融合及び高分子の細胞内輸送を促進し得る。別法では、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)。

【0191】

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。

20

【0192】

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%に医薬的效果がある用量)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。

30

【0193】

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。

40

【0194】

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

50

であろう。

【0195】

(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。

10

【0196】

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。

20

【0197】

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。

【0198】

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。

30

【0199】

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: 40-78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。

40

【0200】

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

50

の放射性核種、或いはアビジン/ビオチン (b i o t i n) 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。

【 0 2 0 1 】

G C R E C をコードするポリヌクレオチド配列を用いて、G C R E C 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 (M C T D)、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、癩癧、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、G e r s t m a n n - S t r a u s s l e r - S c h e i n k e r 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 (c e r e b 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 (S A D) 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 (c o r t i c o b a s a l d e g e n e r a t i o n) 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 (b a l l o o n a n g i o p l a s t y)、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 (m i t r a l a n n u l a r c a l c i f i c a t i o n)、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、腹部アンギナ、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 (A I D S) 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 (A I D S) 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、脾炎、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。MADをコードするポリヌクレオチド配列は、サザーン法やノーザン法、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GCREC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

【0202】

ある実施態様では、GCREC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。GCREC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GCREC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。

【0203】

GCREC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、GCREC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。

【0204】

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。

【0205】

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。

【0206】

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 *i n v i t r o* 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。

【0207】

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 シークエ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 p l i m e r) の検出をすることが可能になる。更に、インシリコ S N P (i n s i l i c o S N P : i s S N P) と呼ばれる配列データベース分析法は、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 D N A 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、D N A 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クエンシングエラーや研究室での D N A 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 M A S S A R R A Y システム (S e q u e n o m , I n c . , S a n D i e g o C A) を用いた質量分析により S N P を検出し、特徴付ける。

【0208】

G C R E C 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 (c o a m p l i f i c a t i o n) 、及び標準的な曲線に結果が加えられたものが含まれる (例えば、M e l b y , P . C . ら (1 9 9 3) J . I m m u n o l . M e t h o d s , 1 5 9 : 2 3 5 - 4 4 ; D u p l a a , C . ら (1 9 9 3) A n a l . B i o c h e m . 2 2 9 - 2 3 6 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。

【0209】

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。

10

20

30

40

50

【0210】

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。

【0211】

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 10
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。

【0212】

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。 20

【0213】

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) 30
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レメント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 40
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。

【0214】

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 50

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。

【0215】

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。

10

20

【0216】

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。

30

【0217】

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。

40

【0218】

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

50

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。

【0219】

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。

10

【0220】

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 Approach, 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。

20

【0221】

本発明の別の実施例ではまた、GCREC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)の遺伝とが関連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)。

30

40

【0222】

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付けることができる。物理的な染色体地図上のGCREC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。

50

【0223】

染色体標本の i n s i t 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るなくても、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、G a t t i , R . A . 他による(1988) *N a t u r e* 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。

10

【0224】

本発明の別の実施例では、G C R E C、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。G C R E C 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。

【0225】

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、G e y s e n , 他による(1984) P C T 出願番号 W O 8 4 / 0 3 5 6 4 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、G C R E C、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G C R E C が、当分野で周知の方法で検出される。精製されたG C R E C 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。

20

【0226】

別の実施例では、G C R E C と結合可能な中和抗体がG C R E C 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、G C R E C 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。

30

【0227】

別の実施例では、発展途上の分子生物学技術にG C R E C 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。

【0228】

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、以下に記載する特定の好適な実施例は、例示目的であって本発明を限定するものではない。

40

【0229】

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60/172,852号、同第60/171,732号、同第60/176,148号、および同第60/177,331号に言及することをもって本明細書の一部とする。

【0230】

(実施例)

1 c D N A ライブラリの作製

インサイトcDNAはL I F E S E Q G O L D データベース (I n c y t e G e n o

50

m i c s , P a l o A l t o C A) に含まれている c D N A ライブラリに由来し、表 4 の列 5 に示されている。S E Q I D N O : 4 0 と示されているインサイト c D N A は、骨髄組織から作製された c D N A ライブラリに由来する。S E Q I D N O : 4 1 のインサイト c D N A は、クローン病に関連する組織を含む小腸組織、大腸組織、および脳組織から作製された c D N A ライブラリに由来する。S E Q I D N O : 4 2 のインサイト c D N A は、前立腺腫瘍組織、小腸組織、乳房組織、および表皮組織から作製された c D N A ライブラリに由来する。S E Q I D N O : 4 3 のインサイト c D N A は、ハンチントン病に関連する脳組織、軟組織腫瘍組織、および胎児肋骨組織から作製された c D N A ライブラリに由来する。S E Q I D N O : 5 7 のインサイト c D N A は、リンパ球および肥満細胞、乳房組織、子宮組織、前立腺組織、副腎組織、脊髄組織、脛骨筋組織、肺組織、食道組織、小腸、および結腸組織から作製された c D N A ライブラリに由来する。S E Q I D N O : 5 8 のインサイト c D N A は、ファロピー管腫瘍組織、子宮内膜組織、および気管支組織から作製された c D N A ライブラリに由来する。S E Q I D N O : 5 9 のインサイト c D N A は、盲腸腫瘍組織を含む結腸組織、膵臓腫瘍組織、脳下垂体組織、および脳組織から作製された c D N A ライブラリに由来する。S E Q I D N O : 6 0 のインサイト c D N A は、脳腫瘍組織およびハンチントン病に関連する組織を含む脳組織、前立腺腫瘍組織、頸部腺癌組織、乳房組織、小腸組織、および膀胱組織から作製された c D N A ライブラリに由来する。まず、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 T R I Z O L (L i f e T e c h n o l o g i e s)、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 R N A を沈殿させた。

10

20

30

40

50

【0231】

R N A の純度を高めるために R N A 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、D N A 分解酵素で R N A 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 (P r o m e g a) または O L I G O T E X ラテックス粒子 (Q I A G E N , V a l e n c i a C A)、O L I G O T E X m R N A 精製キット (Q I A G E N) を用いてポリ (A +) R N A を単離した。別法では、P O L Y (A) P U R E m R N A 精製キット (A m b i o n , A u s t i n T X) などの別の R N A 単離キットを用いて組織溶解物から直接単離した。

【0232】

ある場合には、S t r a t a g e n e 社に R N A を提供し、S t r a t a g e n e 社が対応する c D N A ライブラリを作製した。そうでない場合は、U N I Z A P ベクターシステム (S t r a t a g e n e) または S U P E R S C R I P T プラスミドシステム (L i f e T e c h n o l o g i e s) を用いて当分野で周知の推奨方法または類似の方法で c D N A を合成して c D N A ライブラリを作製した。(例えば、A u s u b e l , 1 9 9 7 , 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 c D N A 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 c D N A を消化した。殆どのライブラリでは、S E P H A C R Y L S 1 0 0 0 または S E P H A R O S E C L 2 B、S E P H A R O S E C L 4 B カラムクロマトグラフィー (A m e r s h a m P h a r m a c i a B i o t e c h)、アガロースゲル電気泳動法によって c D N A の大きさ (3 0 0 ~ 1 0 0 0 b p) を選択した。P B L U E S C R I P T プラスミド (S t r a t a g e n e) または p S P O R T 1 プラスミド (L i f e T e c h n o l o g i e s)、p c D N A 2 . 1 プラスミド (I n v i t r o g e n C a r l s b a d C A)、p I N C Y プラスミド (I n c y t e P h a r m a c e u t i c a l s , P a l o A l t o C A) 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 c D N A を結合させた。この組換えプラスミドを、S t r a t a g e n e 社の X L 1 - B l u e ,

XL1-BIueMRF、SOLR、またはLife Technologies社のDH5 またはDH 10B、ELECTROMAX DH 10Bを含むコンピテント大腸菌細胞に導入し組み込んだ。

【0233】

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(Strata gene) 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 Plus 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。

【0234】

別法では、ハイスルーブットの直接結合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。

【0235】

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ABI CATALYST 800 (PE Biosystems) サーマルサイクラー或いはPTC-200 thermal cycler (MJ Research)などのハイスルーブ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(Molecular Dynamics)、標準ABIプロトコル及び塩基対呼び出しソフトウェアを用いるABI PRISM 373または377シークエンシングシステム(PE Biosystems)、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。cDNA配列内の読み枠は、標準的な方法(Ausubel, 1997, 前出, unit 7.7)を用いて決定した。cDNA配列の幾つかを選択して、実施例8に記載した方法で配列を伸長した。

【0236】

インサイトcDNAに由来する本ポリヌクレオチド配列の確認は、BLAST、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(dinucleotide nearest neighbor analysis)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cDNA配列およびそれらの翻訳を、公共のデータベースである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、およびBLOCKS、PRINTS、DOMO、PRODOM、

およびP F A Mなどの隠れマルコフモデル(HMM)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(HMM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、Eddy, S. R. (1996) Curr. Opin. Struct. Biol. 6: 361-365を参照)。このような問合せは、BLAST、FASTA、BLIMPS、およびHMMRに基づいたプログラムを用いて行った。インサイトcDNA配列を組み立てて、ポリヌクレオチド配列を作製した。或いは、GenBank cDNAs、GenBank EST、ステッチ配列(stitched sequence)、ストレッチ配列(stretched sequences)、またはGenscan-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cDNA群を完全長の配列に伸長した。配列の組み立ては、Phred、Phrap、およびConsedに基づいたプログラムを用いて行い、GeneMark、BLAST、およびFASTAに基づいたプログラムを用いてcDNA群をスクリーニングし、オープンリーディングフレームを決定した。これらのポリヌクレオチド配列を翻訳して対応するポリペプチド配列を得た。次に、これらのポリペプチド配列をGenBankタンパク質データベース(genpept)、SwissProt、BLOCKS、PRINTS、DOMO、PRODOM、Prosite、およびP F A Mなどの隠れマルコフモデル(HMM)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、MACDNAS IS PROソフトウェア(Hitachi Software Engineering, South San Francisco CA)およびLASERGENEソフトウェア(DNA STAR)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MEGALIGNマルチシーケンズアラインメントプログラム(DNA STAR)に組み込まれたCLUSTAL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。

【0237】

表7は、インサイトcDNA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)。

【0238】

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: 40-78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。

【0239】

4 ゲノムDNA由来のコード配列の同定および編集

推定Gタンパク質結合受容体は、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 タンパク質結合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 タンパク質結合受容体について問合せて分析した (7tm__1、7tm__2、7tm__3、および 7tm__4)。潜在的な G タンパク質結合受容体が、G タンパク質結合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 gbprl 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。SEQ ID NO: 40、SEQ ID NO: 41、SEQ ID NO: 42、SEQ ID NO: 43、SEQ ID NO: 61、SEQ ID NO: 62、および SEQ ID NO: 63 を含む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。また、SEQ ID NO: 44、SEQ ID NO: 45、SEQ ID NO: 46、SEQ ID NO: 64、SEQ ID NO: 65、SEQ ID NO: 66、SEQ ID NO: 67、SEQ ID NO: 68、SEQ ID NO: 69、SEQ ID NO: 70、SEQ ID NO: 71、SEQ ID NO: 72、SEQ ID NO: 73、SEQ ID NO: 74、SEQ ID NO: 75、SEQ ID NO: 76、SEQ ID NO: 77、および SEQ ID NO: 78 を含むポリヌクレオチド配列は、その全体が編集された或いは未編集の Genscan 推定コード配列に由来する、完全長コード領域である。また、SEQ ID NO: 47、SEQ ID NO: 48、SEQ ID NO: 49、SEQ ID NO: 50、SEQ ID NO: 51、SEQ ID NO: 52、SEQ ID NO: 53、SEQ ID NO: 54、SEQ ID NO: 55、および SEQ ID NO: 56 を含むポリヌクレオチド配列は、Genscan 推定配列のみを組み立ておよび編集して作製した部分的な遺伝子である。

10

20

30

【0240】

5 ゲノム配列データと cDNA 配列データとの組み立て 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、関連する cDNA および 1 或いは複数のゲノム配列に由来する関連する推定 Genscan エキソンを含む複数のクラスターに入れた。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、後に確認される潜在的なスプライスバリエーションを生成し、編集或いは伸長して完全長の配列を作製した。或るクラスターの 2 つ以上の配列に或る区間の全長が存在する配列区間を同定し、推移 (transitivity) により同定した区間を同等と考える。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、関連しないが連続するゲノム配列を cDNA 配列によって繋ぎ 1 つにする。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で縫い合わせ、可能な最も長い配列および変異配列を作製する。或るタイプ (cDNA と cDNA、またはゲノム配列とゲノム配列) の親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 gbprl 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

40

50

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。

【0241】

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 或いは GenScan エキソン推定配列の何れかと比較した。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 の相同タンパク質に対して、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 の相同タンパク質およびキメラタンパク質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得られたストレッチ配列を検査した。

10

【0242】

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: 40 - 78 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: 40 - 78 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

【0243】

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、1 cM 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

【0244】

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、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)。

【0245】

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

50

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、

【0246】

【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

10

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%の一致で一端が70%重畳しているか、或いは88%一致で他端が100%重畳しているかの何れかの場合である。積スコア50は、100%の一致で一端が50%重畳しているか、或いは79%の一致で他端が100%重畳しているかの何れかの場合である。

20

【0247】

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、脾臓、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

【0248】

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。

40

【0249】

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。

50

【0250】

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ーで増幅を行った。

ステップ1	94	で3分間	10
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ーで増幅を行った。

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	20
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。	

【0251】

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PICOGREEN定量試薬(0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR)を不透明な蛍光光度計プレート(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。

【0252】

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コレラウイルスエンドヌクレアーゼ(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(Stratagene)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/2Xカルベニシリン培養液の384ウェルプレートに37で一晚培養した。

【0253】

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

10

20

30

40

50

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。

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ไซด์(dimethyl sulphoxide)(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。 10

【0254】

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。

【0255】

20

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: 40 - 78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、OLIGO 4.06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 µCi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

【0256】

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-或いは別のイメージ化手段で視覚化して比較する。 40

【0257】

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 50

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 (S ch e n a , M . 他 (1 9 9 5) S c i e n c e 2 7 0 : 4 6 7 - 4 7 0 、 S h a l o n . D . 他 (1 9 9 6) G e n o m e R e s . 6 : 6 3 9 - 6 4 5 、 M a r s h a l l , A . a n d J . H o d g s o n (1 9 9 8) N a t . B i o t e c h n o l . 1 6 : 2 7 - 3 1 . を参照)。

【0258】

完全長cDNA、発現遺伝子配列断片 (E S T)、或いはそれらの断片やオリゴマーが、
10
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、L A S E R G E N Eソフトウェア (D N A S T A R)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
20
調整及び使用について、以下に詳述する。

【0259】

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 (d
T)セルソ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、M M L V逆転写酵素、0.05 pg / μ lのオリゴ (d T)プライマー (21 m
e r)、1 × 第1鎖緩衝液、0.03単位 / μ lのRNアーゼインヒビター、500 μ M
d A T P、500 μ M d G T P、500 μ M d T T P、40 μ M d C T P、40 μ M
d C T P - C y 3 (B D S) または d C T P - C y 5 (A m e r s h a m P h a r m a
c i a B i o t e c h) を用いて逆転写する。この逆転写反応は、G E M B R I G H T
30
キット (I n c y t e) 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、i n v i t r o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C y 3標識、他方はC y 5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RN
Aを変性する。サンプルは、2つの連続するC H R O M A S P I N 30ゲル濾過スピン
カラム (C L O N T E C H L a b o r a t o r i e s , I n c . (C L O N T E C H
) , P a l o A l t o C A) 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10
0%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、S p e e d V A C (S
40
a v a n t I n s t r u m e n t s I n c . , H o l b r o o k N Y) を用いて乾燥
して仕上げ、14 μ l 5 × S S C / 0.2% S D S 中で再懸濁する。

【0260】

マイクロアレイの準備

本発明の配列を用いて、アレイエレメントを作製する。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 μ g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、S E P H A C R Y L - 4 0 0 (A m e r s
h a m P h a r m a c i a B i o t e c h) を用いて精製する。
50

【0261】

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。

【0262】

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。

【0263】

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。

【0264】

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ピ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。

【0265】

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 μmの解像度でスキャンする。

【0266】

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

10

20

30

40

50

された好適なフィルター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、Cy 3では565 nm、Cy 5では650 nm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ーを用いて、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。

【0267】

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較正する。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、較正は2つの蛍光体を有する較正する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。

10

【0268】

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

【0269】

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。

【0270】

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。

30

【0271】

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GCRECを発現する。真核細胞でのGCREC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

40

50

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、GCREC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

【0272】

殆どの発現系では、GCREC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GCREC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GCRECを直接用いて以下の実施例16、17、及び18のアッセイを行うことができる。

20

【0273】

1.3 機能のアッセイ

GCREC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GCREC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、pCMVSPORTTM (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。

30

40

50

【0274】

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

【0275】

14 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。

【0276】

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)。

20

【0277】

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

【0278】

15 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ーにより実質的に精製する。イムノアフィニティークラムは、CNBr-活性化SEPHAROSE(Amersham Pharmacia Biotech)のような活性化クロマトグラフィー用レジンと抗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

【0279】

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、pH2~3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。

【0280】

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

50

などのシグナル伝達物質に關与する分子が含まれる。GCRECまたはその断片を、¹²⁵I ボルトンハンター試薬（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。

【0281】

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)。

【0282】

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。

【0283】

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。この作製物を細胞に形質転換する。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 [³²P]ADP-リボシル化を検出する。SDSサンプルバッファー (4.6% (w/v) SDS、10% (v/v) -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。 [³²P]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法を実施する。これらのゲルをニトロセルロース紙に移し、室温で1時間、blotto (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, Cappe

10

20

30

40

50

l, 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。

【0284】

17 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。

10

【0285】

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)。

20

【0286】

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)。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、1 M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。

30

40

【0287】

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10 mM 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

50

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。

【0288】

18 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Ca²⁺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Ca²⁺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Ca²⁺の動員に關与する経路をGCRECのシグナル伝達を通るように、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Ga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. Thorne (1996) Nature 384 (supp.): 14-16)。受容体を、既知のGPCRリガンド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。

10

20

30

【0289】

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。

【0290】

(表の簡単な説明)

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に対する系統的な名称を示す。

【0291】

表2は、本発明の各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。各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。

【0292】

表3は、推定上のモチーフおよびドメインを含む各ポリペプチド配列の構造的な特徴、並びに各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

40

50

を示す。

【0293】

表4は、各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。

【0294】

表5は、本発明の各ポリヌクレオチドの代表的なcDNAライブラリを示す。

【0295】

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。

【0296】

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、閾値パラメーターを示す。

【0297】

【図面の簡単な説明】

【図1】

MacDNASIS Pro ソフトウェアを用いて作成されたGCREC-1 (SEQ ID NO: 1; インサイト識別番号10491CD1) の疎水性プロットを示す図であって、正のX軸はアミノ酸の位置を示し、負のY軸は疎水性を示し、番号は推定膜貫通ドメインの位置を示す。

10

【図2】

GCREC-3 (SEQ ID NO: 3; インサイト識別番号3168839CD1) の疎水性プロットを示す図。

20

【図3】

GCREC-4 (SEQ ID NO: 4; インサイト識別番号3291235CD1) の疎水性プロットを示す図。

【図4】

GCREC-5 (SEQ ID NO: 5; インサイト識別番号7472001CD1) の疎水性プロットを示す図。

【図5】

GCREC-6 (SEQ ID NO: 6; インサイト識別番号7472003CD1) の疎水性プロットを示す図。

30

【図6】

GCREC-7 (SEQ ID NO: 7; インサイト識別番号7472004CD1) の疎水性プロットを示す図。

【図7】

GCREC-19 (SEQ ID NO: 19; インサイト識別番号3068234CD1) の疎水性プロットを示す図。

【図8】

GCREC-20 (SEQ ID NO: 20; インサイト識別番号5029478CD1) の疎水性プロットを示す図。

40

【図9】

GCREC-21 (SEQ ID NO: 21; インサイト識別番号5102576CD1) の疎水性プロットを示す図。

【表1】

表1-1

インサイトプロジェクトID	ポリペプチドSEQ ID NO	ポリペプチドSEQ ID NO	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
104941	1	104941CD1	40	104941CBI
1499408	2	1499408CD1	41	1499408CBI
3168839	3	3168839CD1	42	3168839CBI
3291235	4	3291235CD1	43	3291235CBI
7472001	5	7472001CD1	44	7472001CBI
7472003	6	7472003CD1	45	7472003CBI
7472004	7	7472004CD1	46	7472004CBI
7475687	8	7475687CPI	47	7475687CBI
7483029	9	7483029CPI	48	7483029CBI
7477933	10	7477933CPI	49	7477933CBI
7475164	11	7475164CPI	50	7475164CBI
7473909	12	7473909CPI	51	7473909CBI
7475252	13	7475252CPI	52	7475252CBI
7927572	14	7927572CPI	53	7927572CBI
7481257	15	7481257CPI	54	7481257CBI
7485790	16	7485790CPI	55	7485790CBI
7482993	17	7482993CPI	56	7482993CBI
2829053	18	2829053CPI	57	2829053CBI
3068234	19	3068234CPI	58	3068234CBI
5029478	20	5029478CPI	59	5029478CBI
5102576	21	5102576CPI	60	5102576CBI
2200534	22	2200534CPI	61	2200534CBI
3275821	23	3275821CPI	62	3275821CBI
3744167	24	3744167CPI	63	3744167CBI
7472007	25	7472007CPI	64	7472007CBI
7472008	26	7472008CPI	65	7472008CBI
7472013	27	7472013CPI	66	7472013CBI
7472015	28	7472015CPI	67	7472015CBI
7472016	29	7472016CPI	68	7472016CBI

10

20

30

40

【表2】

表1-2

インサイト ポリヌクレオチド ID	ポリペプチド 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
7472017	30	7472017CDB1	69	7472017CDB1
7472018	31	7472018CDB1	70	7472018CDB1
7472019	32	7472019CDB1	71	7472019CDB1
7472021	33	7472021CDB1	72	7472021CDB1
7472009	34	7472009CDB1	73	7472009CDB1
7472010	35	7472010CDB1	74	7472010CDB1
7472011	36	7472011CDB1	75	7472011CDB1
7472012	37	7472012CDB1	76	7472012CDB1
7472014	38	7472014CDB1	77	7472014CDB1
7472020	39	7472020CDB1	78	7472020CDB1

10

20

30

40

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 相合体
1	104941CD1	g7211316	6.9E-146	臭菌受容体 [マモセット]
2	1499408CD1	g202806	5.20E-162	パソブレシ受容体 [トブネズミ]
3	3168839CD1	g3618229	2.2E-44	Gタンパク質様P2Y4受容体 [トブネズミ]
4	3291235CD1	g3287369	1.40E-126	A-2 [ハツカネズミ]
5	7472001CD1	g1256393	2.20E-122	味覚受容体タンパク質 TB641 [トブネズミ]
6	7472003CD1	g4378765	1.20E-169	オースティンGタンパク質結合受容体 GPR64 [トブネズミ]
7	7472004CD1	g1698952	6.30E-118	高親和性リノスファチド酸受容体 [アリカツメカエル]
8	7475687CP1	g1256393	4.70E-90	味覚受容体タンパク質 TB641 [トブネズミ]
9	7483029CP1	g2447219	2.50E-75	OLF4 [ヒト]
10	7477933CP1	g2792016	4.90E-79	臭菌受容体 [ヒト]
11	7475164CP1	g517366	3.00E-111	臭菌受容体 [トブネズミ]
12	7473909CP1	g4680264	1.90E-22	臭気物質受容体S55 [ハツカネズミ]
13	7475252CP1	g2447219	6.50E-61	OLF4 [ヒト]
14	7927572CP1	g8100089	6.7E-54	推定味覚受容体HTR2 [ヒト]
15	7481257CP1	g4826521	4.00E-29	αBB481 (新種) 回膜貫通型受容体 (ロトブシニアファミリー) (臭菌受容体様タンパク質 (hsGSM1-15)) [ヒト]
16	7485790CP1	g2447219	3.00E-40	OLF4 [ヒト]
17	7482993CP1	g1314665	3.10E-54	GOLF3 [カニスマリアリス]
18	2829033CD1			
19	3068234CD1	g5922725	3.1E-190	リノスファチン酸Gタンパク質結合受容体 [ヒト]
20	5029478CD1	g1049072	3.6E-21	ガラニン受容体GALRI1 [トブネズミ] (ネズミのRini14B) インスリン/ノルマ細胞からの GALRI1ガラニン受容体のクローニング及び特性。 Brain Res. Mol. Brain Res. 1995年 12月 28日 28:34(2):179-189.)

【表4】

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
21	5102576CD1	g2792016	2.4E-92	臍管受容体 [ヒト] (雄の生殖系列中に発現した臍管受容体遺伝子の分子クローニング及び染色体マッピングヒトゲノムの中にそれらが広く分布していることの詳細, Biochem. Biophys. Res. Commun. 1997年 8月 18日;237(2):283-287.)
22	2200534CD1	g5051404	4.6e-131	537K1.15 (mm17M1-6) 7-膜貫通臍管受容体様タンパク質 (ロドプシンファミリー) [ハツカネズミ]
23	3275821CD1	g182742	1.5e-29	ホリスルベプチド受容体 [ヒト] (Murphy, P.M. 5 (1992) J. Biol. Chem. 267:7637-7643)
24	3744167CD1	g9186902	1.2E-240	ロイコトリエンB4受容体 BLT2 [ハツカネズミ]
25	7472007CD1	g7638409	1.3E-199	臍管受容体P2 [ハツカネズミ]
26	7472008CD1	g4218182	1.0e-89	41271M2.2 (hs6M1-12) (7回膜貫通型受容体 (ロドプシンファミリー) (臍管受容体様) タンパク質) [ヒト]
27	7472013CD1	g205846	2.5e-70	臍管タンパク質 [ドブネズミ]
28	7472015CD1	g1204095	2.5e-25	ドーパミン受容体 [トラフグ]
29	7472016CD1	g6090796	1.1E-215	臍管受容体 [コリウ]
30	7472017CD1	g3757727	2.0e-61	d4801B.7 (臍管受容体様タンパク質 (hs6M1-3)) [ヒト]
31	7472018CD1	g6644328	2.3E-112	オーファンGタンパク質結合受容体OPR26 [ドブネズミ]
32	7472019CD1	g5869916	2.7e-73	臍管受容体 [ハツカネズミ]
33	7472021CD1	g6090804	2.6E-94	臍管受容体 [コリウ]
34	7472009CD1	g1016362	1.6e-68	OL1受容体 [ドブネズミ]
35	7472010CD1	g2317704	7.3e-80	臍管受容体 [ドブネズミ]
36	7472011CD1	g6178008	4.9E-114	臍管タンパク質MOR1B [ハツカネズミ]
37	7472012CD1	g205816	6.8e-84	臍管タンパク質 [ドブネズミ]
38	7472014CD1	g205816	3.9e-88	臍管タンパク質 [ドブネズミ]
39	7472020CD1	g2792016	1.9e-97	臍管受容体 [ヒト]

10

20

30

40

【表 5】

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	104941CD1	311	S68 S189 S292 Y310	N5 N66	7回膜貫通型受容体 (ロドプシンファミリー): G42-Y291 GPCR シグネチャ: K91-P130, I208-Y219, Y283-K299, Y103-S151 臭覚受容体シグネチャ: M60-K81, F178-D192, F239-G254, A275-L286, S292-Q306 膜貫通ドメイン:	MOTIFS HMMEP-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS ProfileScan
2	1499408CD1	891	T148 S686 S114 S248 S350 S481 S501 T628 T814 S856 T84 S140 T144 T325 T411 T543 S568 S676 T706 T788 Y372	N378	ATP/GTP結合部位 (グループ): G202-T209	MOTIFS
3	3168839CD1	422	T232 S178 T342 S363 S371 S397 T21 S211 S226 T307 S332 S367	N4 N9 N251 N323	7回膜貫通型受容体 (ロドプシンファミリー): L39-Y297 ロドプシン様GPCRスーパーファミリー: L24-L48, V57-R78, F101-I23, V137-R158, V192-F215, T232-V256, L279-R305 膜貫通ドメイン: V275-L295	MOTIFS HMMEP-PFAM BLIMPS- BLOCKS ProfileScan BLIMPS- PRINTS HMMEP
4	3291235CD1	609	S228 S229 S396 S456 S324 S328 S364 S417 S466 T506 S568 S590 S153 S268 T392 S462 S482 S560 Y348		7回膜貫通型受容体 (ロドプシンファミリー): E80-E154 GPCRシグネチャ: F76-P115, F395-A405, A442-E458, E509-P526 膜貫通ドメイン: V174-L199	MOTIFS HMMEP-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS HMMEP

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	7472001CD1	313	S68 T194 T200 S267 T309 T138 T164 T290 S306	N5 N85	7回膜貫通型受容体 (ロドプシンファミリー): G41-I259 GPCRシグネチャ: K91-P130, I281-Y297, Y103-A147 臭覚受容体シグネチャ: M60-R81, F178-D192, F239-V254, A273-L284, T290-L304 シグナルペプチド: M1-T38 膜貫通ドメイン: F30-T48, F63-M83	MOTIFS HMNER-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS ProfileScan HMNER SPScan HMNER
6	7472003CD1	398	S36 T155	N10 N18 N28	7回膜貫通型受容体 (ロドプシンファミリー): G59-Y323 GPCRシグネチャ: W108-P147, Y213-Y224, A256-F282, N315-R331, N119-I166 ニューロペプチドY受容体シグネチャ: R89-I81, L321-F334 膜貫通ドメイン: A42-Y65	MOTIFS HMNER-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS ProfileScan HMNER
7	7472004CD1	369	S228 T94 T218 S339 T350	N12	7回膜貫通型受容体 (ロドプシンファミリー): G48-Y321 ロドプシン様GPCRシグネチャ: T33-Y57, I66-F87, F111-I133, R144-V165, V193-L216, A262-V286, S303-I329 膜貫通ドメイン: T33-V51, M109-I125, Y189-W213, W256-V275	MOTIFS HMNER-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS HMNER
8	7475687CP1	194	T186 T76 T82 T46 T172		7回膜貫通型受容体 (ロドプシンファミリー): M1-Y171 オプシナル結合部位: Y142-M194 臭覚受容体シグネチャ: F60-D74, F121-V136, A155-L166, I172-I186	MOTIFS HMNER-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS ProfileScan

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	7483029CPI	173	T16 S34 T60	N32 N167	7回膜貫通型受容体 (ロドプシンファミリー): G8-C146 ロドプシン様GPCRシグネチャ: M26-K47, F71-I93, L107-1128 シグナルペプチド: M1-L22 膜貫通ドメイン: M68-A86, M103-L121	MOTIFS HMER-PFAM BLINPS- BLOCKS ProfileScan BLINPS- PRINTS SPScan HMER
10	7477933CPI	220	S172		7回膜貫通型受容体 (ロドプシンファミリー): P1-Q129 臭覚受容体シグネチャ: M2-K23, F120-D134, F181-G196 ロドプシン様GPCRシグネチャ: L47-I69, I142-V166 膜貫通ドメイン: M44-A62, V85-I110	MOTIFS HMER-PFAM BLINPS- BLOCKS BLINPS- PRINTS HMER
11	7475164CPI	302	T296 S58 S84 T107 T257 T9 T69 S128 T151 S282		7回膜貫通型受容体 (ロドプシンファミリー): G32-I193 GPCRシグネチャ: N81-P120, I273-K289, S93-L142 臭覚受容体シグネチャ: V50-K71, Y168-S182, F229-G244, S265-L276, S282-T296 膜貫通ドメイン: F19-L39, I188-I207	MOTIFS HMER-PFAM BLINPS- BLOCKS ProfileScan BLINPS- PRINTS HMER
12	7473909CPI	110	S70 S36 T66 S94		GPCRシグネチャ: I85-K101 臭覚受容体シグネチャ: F41-G56, A77-L88, S94-Y108	MOTIFS BLINPS- BLOCKS BLINPS- PRINTS

10

20

30

40

【表 8】

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
13	7475252CP1	178	S66 S151 S136	N4 N64	7回膜貫通型受容体 (ロドプシンファミリー): G40-L153 ロドプシン様GPCRシグネチャ: V25-S49, M58-K79, L103-1125, S102-S151 膜貫通ドメイン: L29-145, M100-M117 臭覚受容体シグネチャ: F25-D39	MOTIFS HMME-PFAM BLIMPS- BLOCKS ProfileScan BLIMPS- PRINTS HMME
14	7927572CP1	92			7回膜貫通型受容体 (ロドプシンファミリー): M1-Y96 GPCRシグネチャ: V13-Y24, Q41-Q67 臭覚受容体シグネチャ: F44-G59, L80-L91	MOTIFS HMME-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS
15	7481257CP1	97			シグナルペプチド: M1-G27 膜貫通ドメイン: M8-Y24	SPScan HMME HMME
16	7485790CP1	133	S74		7回膜貫通型受容体 (ロドプシンファミリー): R22-V128 GPCRシグネチャ: G71-P110 膜貫通ドメイン: M62-A100	MOTIFS HMME-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS HMME
17	7482993CP1	213	S85 S205 S159 T183	N83	7回膜貫通型受容体 (ロドプシンファミリー): S2-V182 GPCRシグネチャ: R127-R153, S2-A38 臭覚受容体シグネチャ: F69-N83, F130-G145, V166-L177, T183-G197 膜貫通ドメイン: P102-I120, F130-V152	MOTIFS HMME-PFAM BLIMPS- BLOCKS ProfileScan BLIMPS- PRINTS HMME

10

20

30

40

表 3-5

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
18	2829053CD1	180	S30 S41 S109 S125 S140 S35 S36 S149		β-17ドレナリン作用受容体 シグネチャ: I148-S166 シグナルペプチド: MI-S67	MOTIFS BLIMPS- PRINTS SPScan
19	3068234CD1	353	T146 T217 T233 S321 S17 T21 S294 S329 T141 S229 T303 Y14	N15 N139 N172 N349	7回膜貫通型受容体 (ロドプシンファミリー): S47-Y293 ロドプシン様GPCRスーパーファミリーシグネチャ: I32-I56, F65-L86, L109-I131, R144-L165, Y187-Y210, L237-L261, K276-Y301	MOTIFS HMMER-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS
20	5029478CD1	361	T242 S256 S237 S350	N21 N322	膜貫通ドメイン: I36-I56, T146-G166, Y187-L207, I240-Y258 7回膜貫通型受容体 (ロドプシンファミリー): G57-Y321 ロドプシン様GPCRスーパーファミリーシグネチャ: V42-A66, T74-Y95, M118-I140, A154-V175, Q208-L231, L262-L286, F303-E329	HMMER MOTIFS HMMER-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS
21	5102576CD1	251	S119 S196		膜貫通ドメイン: I45-V65, V124-Q144, V209-I233, L266-N291 7回膜貫通型受容体 (ロドプシンファミリー): R8-C251 GPCRシグネチャ: R57-P96 興奮受容体シグネチャ: M26-K47, L144-D158, F205-G220, A241-C252 膜貫通ドメイン: I66-D68, V109-T134	HMMER MOTIFS HMMER-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS HMMER

10

20

30

40

【表 10】

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
22	2200534C0D1	315	S6 S136 T291	N4 N154	G-タンパク質結合受容体: DM00013(P232750)17-306: S17-L305 DM00013(A570690)15-304: F16-L305 DM00013(P47881)20-308: P20-L305 PD149621: T246-L305 PD000921: C168-L245 PD002495: N4-L47 BL00237: L89-P128, L207-Y218, T282-K298 臭覚受容体 PR000245: M58-P79, F176-G190, V238-G253, V274-L285, T291-L305	HMMER BLIMPS- PRINTS BLIMPS- BLOCKS HMMER-PFAM MOTIFS
23	3275821C0D1	470	T3 T18 T326 T332 T340 S350 S424 S451 T459 S192	N47	G-タンパク質結合受容体: DM00013(P27482)20-317: V34-L306 PD000009: L68-F169 BL00237: W97-P136, G201-H212, A230-A256, N287-R303 GPCRファミリー: F109-V155 ロドプシンGPCRファミリー PR00237: W31-G55, L66-O87, W111-A133, L147-V168, L193-Q216, F235-L259, L277-R303	BLAST- PRODOM BLAST- DOMO HMMER HMMER-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS ProfileScan MOTIFS
					G-タンパク質結合受容体 膜貫通ドメイン: L33-A51, L68-L259	BLAST- PRODOM BLAST- DOMO HMMER

10

20

30

40

【表 1 1】

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
24	3744167CD1	358	T291 S15 T18 S215	N10 N38 N342	G-タンパク質結合受容体: DM00013 (P46092) 27-318; S19-F290 DM00013 (P31391) 41-326; L29-L304 DM00013 (P35414) 22-324; W16-F290 BL00237; W67-P126; F190-Y201, R217-V243, S280-L296 GPCR プロファイル: Y99-V145 ロドプシン GPCR ファミリー PR00237: T22-A46, A57-F78, C101-V123, L137-Y158, L182-L205, Y222-L246, R270-L296	HMMER-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS ProfileScan MOTIFS
					G-タンパク質結合受容体	BLAST- PRODOM BLAST-DBOM
					膜貫通ドメイン: A138-Y159, 637-Y288	HMMER

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
25	7472007CD1	314	S270 S291 S311 T49 S67 S193	N4 N65	シグナルペプチド: M1-Q56 膜貫通ドメイン: L28-V48, L208-M228 7回膜貫通型受容体 (ロドプシンファミリー) シグネチャ: G41-Y280 G-タンパク質結合受容体シグネチャ: 190-F128, 1207-I218, 1282-K298 G-タンパク質結合受容体シグネチャ: Y102-A147 G-タンパク質結合受容体シグネチャ: A110-A125 臭覚受容体シグネチャ: M59-K80, F177-D191, F238-S253, I274-L285, S291-I305 メラノコルチン受容体ファミリー: A5-L63 ロドプシン様GPCRスーパーファミリー: L26-I50, M59-K80, F104-I126, F153-V174, A199-L222, A237-R261, K272-K298 G-タンパク質結合受容体: DM00013 (P23270), 18-311; L23-H306 G-タンパク質結合受容体: DM00013 (P23267) 20-309; L27-I305 G-タンパク質結合受容体: DM00013 (P23275) 17-306; L23-I305 G-タンパク質結合受容体: DM00013 (P30953) 18-306; L19-H306 臭覚受容体 PD000921: F168-L246 臭覚受容体 PD149621: V247-R307	SPScan HMME HMME-PPAM BLIMPS-BLOCKS ProfileScan MOTIFS BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-DOMO BLAST-DOMO BLAST-DOMO BLAST-DOMO BLAST-PRODOM BLAST-PRODOM

10

20

30

40

表 3-9

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
26	7472008CD1	365	S78, T192, S199, T320, S343, S47, S66, S78, S96, S217, T222, T337, T361	N94	<p>膜貫通ドメイン: L226-L244</p> <p>7回膜貫通型受容体 (ロドプシンファミリ-) シグネチャ: 670-7319</p> <p>G-タンパク質結合受容体シグネチャ: K119-P156, L236-S247, K284-Q290, T311-H327</p> <p>臭覚受容体シグネチャ: M88-Q109, V206-Q220, F267-Q282, L303-L314, T320-K324</p> <p>メラノコルチン受容体ファミリ-: V73-L84, L60-L92</p> <p>ロドプシン様GPCRスーパーファミリ-: L55-L79, M88-Q109, F133-Y155, W228-A251, A266-Q290, K301-H327</p> <p>臭覚受容体 PD000921: L195-L276</p> <p>臭覚受容体 PD149621: V276-K331</p> <p>G-タンパク質結合受容体: DM00013(P30955)18-305; L61-L326</p> <p>G-タンパク質結合受容体: DM00013(P23269)15-304; L61-L326</p> <p>G-タンパク質結合受容体: DM00013(A57069)15-304; D59-L326</p> <p>G-タンパク質結合受容体: DM00013(P23275)17-306; T67-L326</p>	<p>HMMER</p> <p>HMMER-PFAM</p> <p>BLIMPS-BLOCKS</p> <p>BLIMPS-PRINTS</p> <p>BLIMPS-PRINTS</p> <p>BLIMPS-PRINTS</p> <p>BLAST-PRODOM</p> <p>BLAST-PRODOM</p> <p>BLAST-DMO</p> <p>BLAST-DMO</p> <p>BLAST-DMO</p> <p>BLAST-DMO</p>

10

20

30

40

【表 1 4】

表3-10

SEO ID NO:	インサイトボリスチド 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
27	7472013CD1	317	S68 S193	N6 N22 N43	シグナルペプチド: 非-G42 膜貫通ドメイン: I23-L41 7回膜貫通型受容体 (ロドプシンファミリー)シグネチャ: G42-L155, A279-Y295 G-タンパク質結合受容体シグネチャ: P91-F130, N212-Y223, T287-K303 G-タンパク質結合受容体シグネチャ: F163-L148 G-タンパク質結合受容体シグネチャ: T111-A126 臭受容体シグネチャ: M60-Q81, F182-D196, V243-G258, A279-A290, S298-L310 メラノコルチン受容体ファミリー: M52-L84 ロドプシン様GPCRスーパーファミリー: L27-S51, M60-Q81, F105-I127, R141-G162, I204-G221, A242-G286, M277-L303 GPCRスーパーファミリー受容体シグネチャ: V161-M178 G-タンパク質結合受容体: DM00013 (P23274) 18-306; L27-L310 G-タンパク質結合受容体: DM00013 (P23272) 18-306; Y25-L310 G-タンパク質結合受容体: DM00013 (P30953) 18-306; L27-L310 G-タンパク質結合受容体: DM00013 (P30955) 18-306; L27-L310 臭受容体 P0000921: G174-L250 臭受容体 PD149621: T251-L310	SPScan HMMER HMMER-PFAM BLIMPS-BLOCKS ProfileScan MOTIFS BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-DOMO BLAST-DOMO BLAST-DOMO BLAST-PRODOM BLAST-PRODOM

【表 15】

10

20

30

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, ドメイン及びモチーフ	分析方法及びデータベース
28	7472015CD1	335	T73 S79 S214 S309 T217 S329 S331		シグナルペプチド: M1-420 シグナルペプチド: M1-420 膜貫通ドメイン: F9-V27, L45-I63, M117-I136 7回膜貫通受容体 (ロドプシンファミリー) シグネチャ: I271-I279	HMMER SPScan HMMER HMMER-PPAM
					G-タンパク質結合受容体シグネチャ: R71-P110, F174-Y186, P218-T244, N271-R287	BLIMPS- BLOCKS
					G-タンパク質結合受容体シグネチャ: F84-L129	ProfileScan
					G-タンパク質結合受容体シグネチャ: F91-I106	MOTIFS
					ロドプシン様ドメインスーパーファミリー: S6-L30, S40-L61, V85-I102, E21-R142, V166-L189, A223-V247, E281-R287	BLIMPS- PRINTS
					G-タンパク質結合受容体: DM00013 (P41596) 137-461; G8-P220	BLAST-DOMO
					G-タンパク質結合受容体: DM00013 (P47800) 29-538; G8-Y281	BLAST-DOMO
					G-タンパク質結合受容体: DM00013 (P31388) 20-336; G8-P218	BLAST-DOMO
					G-タンパク質結合受容体: DM00013 (JN0591) 20-336; G8-P218	BLAST-DOMO

10

20

30

40

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
29	7472016CD1	309	S8, S67, S188, S266, S137, S229, S266, S289	N5, N65, N264	シグナルペプチド: M1-L55 膜貫通ドメイン: I28-A46, M199-I218 7回膜貫通型受容体 (ロドプシンファミリー)シグネチャ: G41-I288 G-タンパク質結合受容体シグネチャ: N30-P129, F209-L296 L294-A280, I280-K296 G-タンパク質結合受容体シグネチャ: Y102-A148 異源受容体シグネチャ: R69-A80, F177-D191, F237-G252, I272-L283, S289-F303 ロドプシン様GPCRスーパーファミリー: F26-G50, R59-A80, M104-I105, S140-L161, M188-F219, A270-K296 G-タンパク質結合受容体: DM00013 (P23256) I7-306; 020-L302 G-タンパク質結合受容体: DM00013 (P23274) S-306; E22-L299 G-タンパク質結合受容体: DM00013 (P23269) I5-304; 021-L299 G-タンパク質結合受容体: DM00013 (P30935) S-305; E22-L302 異源受容体 PD148621: I245-S309 異源受容体 PD000921: L106-L244	SPScan HMMER HMMER-PPAM BLIMPS- BLOCKS ProfileScan BLIMPS- PRINTS BLIMPS- PRINTS BLAST-DOMO BLAST-DOMO BLAST-DOMO BLAST-DOMO BLAST- PRODOM BLAST- PRODOM

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
30	7472017CD1	236	S7 T217	N5 N189	シグナルペプチド: M1-G42 機道ドメイン: G31-H52, V123-L141 7回膜貫通型受容体 (ロドシリンファミリー)シグネチャ: F12-Y216 G-タンパク質結合受容体シグネチャ: K24-P63, L133-V144, G161-Q187, T208-K224 G-タンパク質結合受容体シグネチャ: Y36-Y81 G-タンパク質結合受容体シグネチャ: I44-A89 異質受容体シグネチャ: L164-G179, I200-L211, T217-N231 ロドシリン様GPCRスーパーファミリー: S38-V63, L125-A148, G163-Q187, K198-K224 異質受容体 PD148621: V173-I236 異質受容体 PD000921: G163-I172 G-タンパク質結合受容体: DM00013 (P23269)15-304: L15-L227 G-タンパク質結合受容体: DM00013 (P20953)18-306: L15-L227 G-タンパク質結合受容体: DM00013 (A57069)15-304: L15-R228 G-タンパク質結合受容体: DM00013 (P23275)17-306: M1-L227	SPScan HMMER HMMER-PFAM BLIMPS- BLOCKS ProfileScan MOTIFS BLIMPS- PRINTS BLIMPS- PRINTS BLAST- PRODOM BLAST- PRODOM BLAST-DOMO BLAST-DOMO BLAST-DOMO BLAST-DOMO

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	潜在的リジン酸化部位	潜在的 クロニコシル化部位	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
31	7472018CD1	363	Y294 S321 S325 T353 S157 T210 S223 T240 T316 T340	N47 N348 N355	シグナルペプチド: M1-A24 シグナルペプチド: M1-A24 7回膜貫通型受容体 (ロドプシンファミリー)シグネチャ: S22-Y294 G-タンパク質結合受容体シグネチャ: F72-P111, F181-S192, R234-T260, K286-R302 ロドプシン様GPCRスーパーファミリー: L7-A31, S41-F62, D86-V108, Y122-G143, T173-H198, A238-A263, G276-R302 P2Y4プリン受容体シグネチャ: Y32-L48, P111-L126 G-タンパク質結合受容体: DM00013 (P50931)20-336; P3-L305 G-タンパク質結合受容体: DM00013 (P53452)17-344; L7-F268 G-タンパク質結合受容体: DM00013 (P50408)20-335; 64-L305 G-タンパク質結合受容体: DM00013 (P31388)20-336; P3-L305	HMMER SPScan HMMER-PPAM BLIMPS- BLOCKS BLIMPS- PRINTS BLIMPS- PRINTS BLAST-DOMO BLAST-DOMO BLAST-DOMO BLAST-DOMO

10

20

30

40

【表 19】

表3-15

SEQ ID NO.	インサート 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
32	7472019CD1	308	S162 S290 S67 T187 S192 S265	N5 N65	腫瘍ドメイン: L30-149, M197-L215 7回膜貫通型受容体 (ロドプシンファミリー) シグネチャ: 641-Y289 G-タンパク質結合受容体シグネチャ: P90-P129, L206-Y217, L234-K280, T281-K297 G-タンパク質結合受容体シグネチャ: S102-T147 臭覚受容体シグネチャ: I69-Q80, F176-D190, F237-G252, L273-L284, S290-M304 メラノコルチン受容体ファミリー: F61-L63 ロドプシン様GPCRスーパーファミリー: T26-M50, I59-Q80, S104-L126, V140-L161, I198-K221, K271-K297 G-タンパク質結合受容体: DM00013 (P23275) 17-306; I17-M304 G-タンパク質結合受容体: DM00013 (P23265) 15-304; F27-M304 G-タンパク質結合受容体: DM00013 (P23286) 17-306; I17-M304 G-タンパク質結合受容体: DM00013 (S29707) 8-306; P21-I300 臭覚受容体 PD000921: L165-L244 臭覚受容体 PD149621: T245-M304	HMMER HMMER-PFAM BLIMPS- BLOCKS ProfileScan BLIMPS- PRINTS BLIMPS- PRINTS BLIMPS- PRINTS BLAST-DOMO BLAST-DOMO BLAST-DOMO BLAST-DOMO BLAST- PRODOM BLAST- PRODOM

10

20

30

40

表3-16

SEQ ID NO.	インサイト ポリヌクレオチド 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
33	7472021CD1	343	S87 T154 S288 S326 S311 S316	N25 N183 N314	既述ドメイン: Y65-L75, I214-I234 7回膜貫通型受容体 (ロドプシンファミリー)シグネチャ: G61-Y310 G-タンパク質結合受容体シグネチャ: S38-L64, G110-P149, P302-K318 G-タンパク質結合受容体シグネチャ: F122-V166 G-タンパク質結合受容体シグネチャ: S130-A145 異常受容体シグネチャ: M9-G100, F197-Y211, F258-S273, F297-L305, S311-L325 メラノコルチン受容体ファミリー: V71-L83 パソプレニン受容体シグネチャ: L79-L86 異常受容体 PD000921: I186-L265 異常受容体 PD149621: V267-E328 G-タンパク質結合受容体: DM00013 (P23259) 15-304: E40-L325 G-タンパク質結合受容体: DM00013 (P23275) 17-306: S38-L325 G-タンパク質結合受容体: DM00013 (P23273) 18-306: 145-L325 G-タンパク質結合受容体: DM00013 (P23266) 17-306: S38-S326	HMMER HMMER-PFAM BLIMPS-BLOCKS ProfileScan MOTIFS BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS PRODOM PRODOM BLAST-DOMO BLAST-DOMO BLAST-DOMO BLAST-DOMO

10

20

30

40

表3-17

SEQ ID NO.	インサイトホリベプチド 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
34	7472009CDI	323	S87 S232 T290 S8 S67 T193	N5	膜貫通ドメイン: L30-L47, L201-L221 7回膜貫通型受容体 (ロドプシンファミリー) シグネチャ: G41-Y289 G-タンパク質結合受容体シグネチャ: K90-P129, L206-Y217, R234-R260, T281-A287 G-タンパク質結合受容体シグネチャ: F102-N147 G-タンパク質結合受容体シグネチャ: A110-A126 臭覚受容体シグネチャ: M69-K80, F177-D191, F237-Q252, G273-L284, T290-L304 ロドプシン様GPCRスーパーファミリー: F26-Q50, Y104-I126, V140-A161, T198-L221, K271-A287 メラノコルチン受容体ファミリー: I51-L63, I126-N137 パンプレリン受容体シグネチャ: L55-L66 G-タンパク質結合受容体: DM00013 (P23275) 17-306, L25-L304 G-タンパク質結合受容体: DM00013 (A57069) 15-304, L27-L304 G-タンパク質結合受容体: DM00013 (P23270) 18-311, L25-L304 G-タンパク質結合受容体: DM00013 (P23266) 17-306, L27-L304 臭覚受容体 PD149621: T245-T310 臭覚受容体 PD000921: F168-L244	HMMER HMMER-PFAM BLIMPS-BLOCKS ProfileScan MOTIFS BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-DOMO BLAST-DOMO BLAST-DOMO BLAST-DOMO BLAST-PRODOM BLAST-PRODOM

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
35	7472010CDI	299	T158 S126 S280 T293 S10 S57 T156	N55	シグネチャ (signature) 配列, ドメイン及びモチーフ 膜貫通ドメイン: L188-1206 7回膜貫通型受容体 (ロドプシンファミリー) シグネチャ: G31-1219 G-タンパク質結合受容体シグネチャ: S79-P118, F188-S199, S224-T250, V271-K287 G-タンパク質結合受容体シグネチャ: F91-F135 G-タンパク質結合受容体シグネチャ: S99-A114 鼻受容体シグネチャ: M43-K70, Y168-S180, F227-G242, A269-L274, S280-L294 メラノコルチン受容体ファミリー: I41-L63 ロドプシン様GPCRスーパーファミリー: R1-G40, M19-L30, L15, I181-K204, A226-L236, R261-K287 G-タンパク質結合受容体: DM00013 (S29703) 11-299: 623-L294 G-タンパク質結合受容体: DM00013 (S61356) 18-307: 124-K282 G-タンパク質結合受容体: DM00013 (P23274) 18-306: 124-L294 G-タンパク質結合受容体: DM00013 (P30956) 18-305: 124-L294 鼻受容体 PD149621: V237-K286 鼻受容体 PD000921: L156-1236	HMMER HMMER-PFAM BLIMPS-BLOCKS ProfileScan MOTIFS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-DOMO BLAST-DOMO BLAST-DOMO BLAST-DOMO BLAST-PRODOM BLAST-PRODOM

10

20

30

40

表3-19

SEQ ID NO.	インサイトポリベプチド 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
36	74720.IICD1	307	S87 T288 S193	N5	膜貫通ドメイン: L23-143, M89-M118, G204-H228 7回膜貫通型受容体 (ロドプシンファミリー) シグネチャ: G41-Y287 G-タンパク質結合受容体シグネチャ: K90-P129, R234-R260, T279-Q295 G-タンパク質結合受容体シグネチャ: F102-I146 興奮受容体シグネチャ: M59-K80, F177-D191, A237-V252, V271-L282, T288-6302 メラノコルチン受容体ファミリー: S31-L63 ロドプシン様GPCRスーパーファミリー: F26-T50, M59-K80, F104-I126, L140-A161, K199-L222, A236-R260, K269-Q295 興奮受容体 PD000921: L166-I245 興奮受容体 PD149621: V246-R303 G-タンパク質結合受容体: DM00013 (S2971D) 15-301: L17-L301 G-タンパク質結合受容体: DM00013 (P2321S) 7-306: L17-L301 G-タンパク質結合受容体: DM00013 (P23266) 17-308: L17-L301 G-タンパク質結合受容体: DM00013 (P4788) 12P-309: L23-L301	HMMER HMMER-PFAM BLIMPS-BLOCKS ProfileScan BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODOM BLAST-PRODOM BLAST-DOMO BLAST-DOMO BLAST-DOMO BLAST-DOMO

10

20

30

40

【表 2 4】

表3-20

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
37	7472012CD1	314	T19 S230 S291	潜在的グリコシル化部位 NS N38	シグネチャ (signature) 配列, ドメイン及びモチーフ: 1回糖基遷移受容体 (ロドプシンファミリー) シグネチャ: G41-I230 G-タンパク質結合受容体シグネチャ: K90-P129, I282-K298 G-タンパク質結合受容体シグネチャ: Y102-M147 G-タンパク質結合受容体シグネチャ: M110-A125 興奮受容体シグネチャ: M59-K90, F177-S191, F238-Q253, I274-L285, S291-M305 ロドプシン様GPCRスーパーファミリー: P28-L50, M87-L300, G104-L26, L199-L222, I237-R261, R272-K298 G-タンパク質結合受容体: DM00013 (P23268) I17-306, I117-K303 G-タンパク質結合受容体: DM00013 (P23274) I18-306, E22-K303 G-タンパク質結合受容体: DM00013 (S29107) I18-306, P21-G299 G-タンパク質結合受容体: DM00013 (P30955) I18-306, P21-K303 興奮受容体 PD149821: I246-L310 興奮受容体 PD000921: L166-L245	HMNER-PFAM BLIMPS-BLOCKS ProfileScan MOTIFS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-DOMO BLAST-DOMO BLAST-DOMO BLAST-DOMO BLAST-PRODOM BLAST-PRODOM

10

20

30

40

表3-21

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
38	7472014CD1	310	S19 S67 S93 T267 S18 S137 S290	N5 N265	膜貫通ドメイン: V30-146, M69-178 7回膜貫通型受容体 (ロドプシンファミリー) シグネチャ: G41-7289 G-タンパク質結合受容体シグネチャ: K90-P129, I207-Y218, R235-6261, T281-K297 G-タンパク質結合受容体シグネチャ: Y102-1151 G-タンパク質結合受容体シグネチャ: I110-A1125 膜貫通型シグネチャ: M59-K80, F177-S191, F238-6253, A275-L284, S290-R304 メラノコリン受容体ファミリー: S51-L63 ロドプシン様GPRスーパーファミリー: P25-L50, M59-K80, F104-I126, V199-L222, Q271-K297 鼻受容体 PB000921: L186-L245 鼻受容体 PD149621: T246-R306 G-タンパク質結合受容体: DM00013 (P23276) 7-306: L17-H304 G-タンパク質結合受容体: DM00013 (P23274) 8-306: E22-H304 G-タンパク質結合受容体: DM00013 (P30955) 8-305: D23-H304 G-タンパク質結合受容体: DM00013 (P30953) 8-306: R20-H305	HMMER HMMER-PFAM BLIMPS- BLOCKS ProfileScan MOTIFS BLIMPS- PRINTS BLIMPS- PRINTS BLIMPS- PRINTS BLAST- PRODOM BLAST- PRODOM BLAST-DOMO BLAST-DOMO BLAST-DOMO BLAST-DOMO

10

20

30

40

表3-22

SEQ ID NO.	インサイトポリベプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的シル化部位	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
39	7472020CD1	359	S257 S317 S178 S255	N31	<p>腫瘍ドメイン: M127-A145, V168-T193</p> <p>7回膜貫通型受容体 (ロトリンファファミリー) シグネチャ: R67-Y316</p> <p>G-タンパク質結合受容体シグネチャ: R116-P155, G233-Y244, S261-T287, T308-Q324</p> <p>G-タンパク質結合受容体シグネチャ: F129-V173</p> <p>興奮受容体シグネチャ: M85-K106, F203-D217, F264-G279, A300-L311, S317-R331</p> <p>GPRオプティン受容体シグネチャ: S317-R326</p> <p>カンナビノイド受容体シグネチャ: M60-L73, Y316-A326</p> <p>G-タンパク質結合受容体: DM0013 (P23268) 17-306; E45-L327</p> <p>G-タンパク質結合受容体: DM00013 (P23268) 18-307; S44-L330</p> <p>G-タンパク質結合受容体: DM00013 (P23707) 18-306; P47-L327</p> <p>G-タンパク質結合受容体: DM00013 (P30953) 18-306; P47-L330</p> <p>興奮受容体 PD000921: N197-L271</p> <p>興奮受容体 PD149621: V273-R333</p>	<p>HMMER</p> <p>HMMER-PFAM</p> <p>BLIMPS-BLOCKS</p> <p>ProfileScan</p> <p>BLIMPS-PRINTS</p> <p>BLIMPS-PRINTS</p> <p>BLIMPS-PRINTS</p> <p>BLAST-DOMO</p> <p>BLAST-DOMO</p> <p>BLAST-DOMO</p> <p>BLAST-DOMO</p> <p>BLAST-PRODOM</p> <p>BLAST-PRODOM</p>

10

20

30

40

【表 27】

表4-2

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
57	2829053CB1	1370		170756F1 (EMARNO02)	1	534
				2829053F6 (CLYMN0T03)	428	989
				6098294H1 (UTREN0T09)	849	1143
				5279076H1 (MUSLN0T01)	1071	1308
58	3068234CB1	1567		4588915H1 (MASTXT01)	1132	1370
				70489898V1	1	459
				70488597V1	353	883
				5837294H1 (FTUBT01)	730	983
59	5029478CB1	1321		70490272V1	955	1567
				6035153H1 (PITVNO06)	1	582
60	5102576CB1	1110		6558521H1 (BRAFN002)	504	1190
				3076961F6 (COLCT0T03)	742	1321
				5496406H1 (BRABDIR01)	1	250
				1720010F6 (ELADN0T06)	151	708
61	2200534CB1	1095	1037-1095, 372-491	8969401U1	602	710
				5102576F6 (PROSTUS20)	673	1110
				92905881.V113.gs.2	1	948
				2200534F6 (SPLAFET02)	534	1095
62	3275821CB1	1665	1431-1665, 765-1294, 240-597	576308R6 (BRAVXT04)	490	1051
				3275821F6 (PROSPT06)	1	548
63	3744167CB1	1609	1184-1238, 249-522	93779013.V113.gs.9	265	1665
				2762336H1 (BRSTN0T12)	745	994
				95578767.V113.gs.4	262	1338
				3744167H1 (THYMN0T08)	693	977
64	7472007CB1	945		8835247	1212	1609
				34745866H1 (LUNGN0T27)	1	309
				92431610.V113.gs.4.nt	1	945
				63093312.V113.gs.10.nt	1	1098
65	7472008CB1	1098		94190944.V113.gs.3.nt	1	954
				94467309.V113.gs.2.nt	1	1008
66	7472013CB1	954		94567182.V113.gs.19.nt	1	930
				7472015CB1	1008	
67	7472015CB1	1008		65262456.V113.gs.7.nt	1	711
				7472017CB1	711	
68	7472016CB1	930		65523795.V113.gs.12.nt	1	1092
				7472018CB1	1092	

10

20

30

40

【表 2 9】

表4-3

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
71	7472019CB1	927		95566548.v113.gs_7.nt	1	927
72	7472021CB1	1032		95708153.v113.gs_6.nt	1	1032
73	7472009CB1	972		93213020.v113.gs_4.nt	1	972
74	7472010CB1	900		93738097.v113.gs_9.nt	1	900
75	7472011CB1	924		93924656.v113.gs_5.nt	1	924
76	7472012CB1	945		94190944.v113.gs_1.nt	1	945
77	7472014CB1	933		94190944.v113.gs_4.nt	1	933
78	7472020CB1	1080		95706779.v113.gs_3.nt	1	1080

10

20

30

40

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
61	2200534CB1	BRVITXT04
62	3275821CB1	PROSBPT06
63	3744167CB1	LUNGNOT27

10

20

30

40

表6

ライブラリ	ベクター	ライブラリの説明
BRAVXT04	PSPORT1	ライブラリは、IL-1を含むサイトカインとの組み合わせで4時間から6時間の間刺激されたヒトの星状細胞の分離した集団から単離したRNAを用いて作製した。このRNAは、ポリA RNAの単離及びライブラリの作製のためにプールされていたものである。
LUNGNOT27	P INCY	ライブラリは、17歳のヒスパニック系女性から採取した肺組織から単離したRNAを用いて作製した。
PROSBPT06	p INCY	ライブラリは、66歳の白人男性から根治的前立腺及びリンパ節切除の際に採取した病変前立腺組織から単離したRNAを用いて作製した。病理学的には、腺線維筋腫性過形成を示していた。関連する腫瘍組織は、病理学的には(グレード4中の)グレード2腺癌、Gleason グレード 3+3 を示していた。患者は前立腺特異性抗原(PSA)、タンパク尿、腎機能低下、及び尿頻度を示していた。患者の病歴には、半側不全麻痺、抑うつ障害、睡眠無呼吸、乾癩、僧帽弁逸脱症候群、脳血管障害、良性高血圧症、及びインフルエンザが含まれる。家族歴には、良性高血圧症、脳血管障害、及び大腸癌が含まれる。

10

20

30

表7-1

プログラム名	説明	引用文献	パラメーター閾値
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=1.06E-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、構造的フィングアンプリント領域を探索する。	Henikoff, S and J.G. Henikoff, Nuci, Acid Res., 19:6565-72, 1991. J.C.; Henikoff 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 (1998) 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

【 表 3 3 】

10

20

30

表7-2

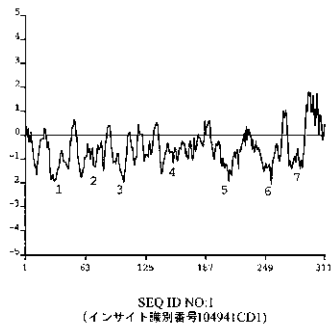
プログラム名	説明	引用文献	パラメーター閾値
ProfileScan	Prosite で定義された配列パターンと一致するタンパク質配列における構造及び配列のモチーフを検索するアルゴリズム。	Gribskov, M. 他 (1988) CABIOS 4:61-66; Gribskov, 他 (1989) Methods Enzymol. 183:146-159; Bairoch, A. 他 (1997) Nucleic Acids Res. 25:217-221	標準化された質のスコア≧特定の Prositeモチーフに対する GCG 指定 "HIGH" 値 通常、スコア=1.4-2.1
Phred	高い感度及び確率で自動配列決定機のアナログ出力を調べる塩基誤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レメンテーションに基づく SWAT や CrossMatch を含む Philis 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=66 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ャン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	タンパク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	タンパク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	Prosite 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.	

10

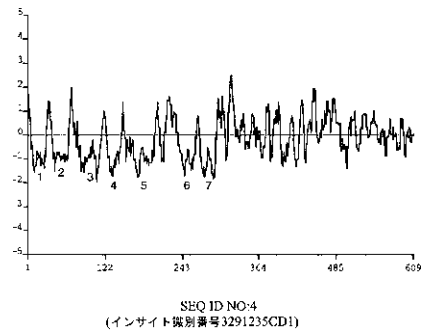
20

30

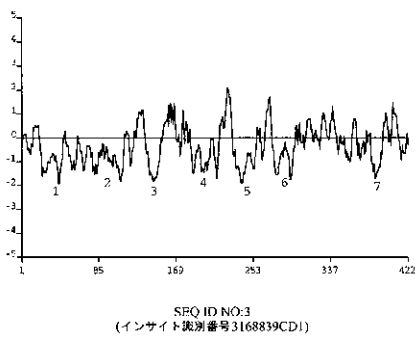
【 図 1 】



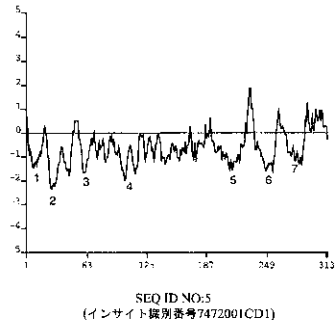
【 図 3 】



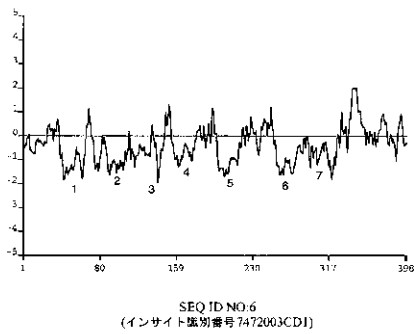
【 図 2 】



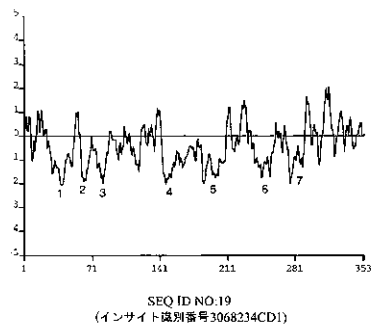
【 図 4 】



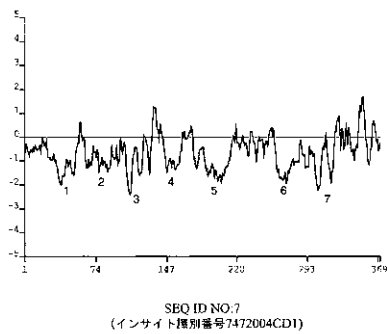
【 図 5 】



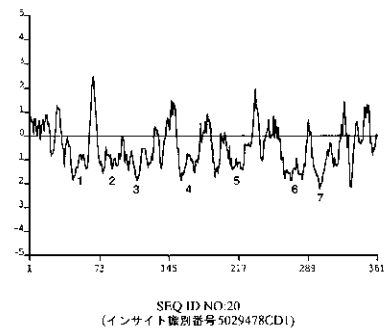
【 図 7 】



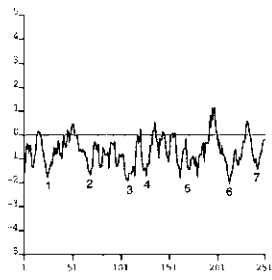
【 図 6 】



【 図 8 】



【 図 9 】



SFQ ID NO:21
(インサイト識別番号S102576CD1)

【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
14 June 2004 (14.06.2004)

PCT

(10) International Publication Number
WO 01/42288 A2

(51) International Patent Classification: C07K 14/00
(21) International Application Number: PCT/US00/33382
(22) International Filing Date: 7 December 2000 (07.12.2000)
(25) Filing Language: English
(26) Publication Language: English

BAUGHN, Mariah, R. (US/US); 14214 Santiago Road, San Leandro, CA 94577 (US); AI-YOUNG, Jaihee (US/US); 233 Golden Eagle Lane, Brisbane, CA 94005 (US); YANG, Junming (CN/US); 7125 Bark Lane, San Jose, CA 95129 (US); LU, Dyung, Aijun, M. (US/US); 233 Coy Drive, San Jose, CA 95123 (US); REDDY, Roopa (IN/US); 1233 W. McKinley Avenue #3, Sunnyvale, CA 94086 (US).

(30) Priority Data:
60/172,852 10 December 1999 (10.12.1999) US
60/171,732 22 December 1999 (22.12.1999) US
60/176,148 14 January 2000 (14.01.2000) US
60/177,531 21 January 2000 (21.01.2000) US

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

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

(81) Designated States (national): AF, AG, AI, AM, AN, 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, LY, MA, MG, MK, MN, MW, MX, MY, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TH, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(72) Inventors and
(75) Inventors/Applicants (for US only): BURFORD, Neil (GB/US); 105 Wildwood Circle, Durham, CT 06422 (US).

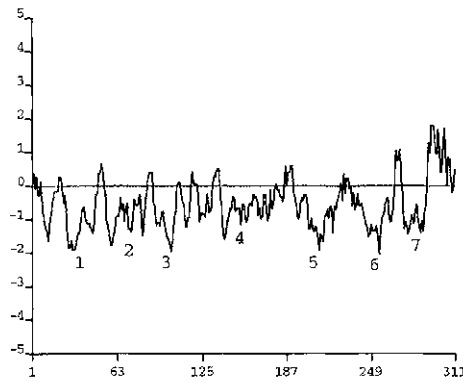
(84) Designated States (regional): ARIPO patent (GH, GM, KI, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW); Eurasian

[Continued on next page]

(54) Title: G PROTEIN COUPLED RECEPTORS



WO 01/42288 A2



(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.

WO 01/42288 A2



patent (AM, AZ, BY, KG, KZ, MD, RU, UJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IL, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

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

WO 01/42288

PCT/US00/33382

G-PROTEIN COUPLED RECEPTORS**TECHNICAL FIELD**

5 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.

10 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.

20 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 01/42288

PCT/US00/33382

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. Arkinstall (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 prostanooids, 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 01/42288

PCT/US00/33382

methyl-D-aspartate/NMDA receptors). (Reviewed in Watson, S. and S. Arkininstall (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
lysophospholipids, which appear to have roles in development and neuropathology (Chun, J. et al.
10 (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
35 (1998) J. Leukoc. Biol. 63:271-280).

WO 01/42288

PCT/US00/33382

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_B receptors, and the taste receptors.

Other subfamilies of GPCRs include two groups of chemoreceptor genes found in the nematodes *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, hypercalcemia); parathyroid hormone (short limbed dwarfism); β_2 -adrenoceptor (obesity, non-insulin-dependent diabetes mellitus); growth hormone releasing hormone (dwarfism); and adrenocorticotropin (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 adrenoceptors have been used for the treatment of asthma, high blood pressure, other cardiovascular

WO 01/42288

PCT/US00/33382

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*).

Recent research suggests potential future therapeutic uses for GPCRs in the treatment of
5 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 (Schöneberg, T. et al. (1996) EMBO J. 15:1283-1291). Mutations in melanocortin-4 receptor (MC4R) are implicated in human weight regulation and obesity.
10 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 1 receptor for parathyroid hormone (PTH) is a GPCR that mediates the PTH-dependent regulation of calcium homeostasis in the bloodstream. Study of PTH/receptor interactions may enable the development of novel PTH receptor ligands for the
15 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 trafficking, hematopoiesis, and angiogenesis. Targeted disruption of various chemokine receptors in
20 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, results in resistance to AIDS, suggesting that CCR5 antagonists could be useful in preventing the
25 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, autoimmune/inflammatory, and metabolic disorders, and viral infections, and in the assessment of the
30 effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of G-protein coupled receptors.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, G-protein coupled receptors, referred to

WO 01/42288

PCT/US00/33382

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," "GCREC-21," "GCREC-22," "GCREC-23," "GCREC-24,"

5 "GCREC-25," "GCREC-26," "GCREC-27," "GCREC-28," "GCREC-29," "GCREC-30," "GCREC-31," "GCREC-32," "GCREC-33," "GCREC-34," "GCREC-35," "GCREC-36," "GCREC-37," "GCREC-38," and "GCREC-39." 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:1-39, b) a naturally occurring amino acid sequence

10 having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-39.

15 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:1-39, 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:1-39, c) a biologically active fragment of an amino acid sequence selected from the group consisting of

20 SEQ ID NO:1-39, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-39. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:40-78.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter

25 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:1-39, 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:1-39, c)

30 a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39. 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

WO 01/42288

PCT/US00/33382

sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, 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:1-39, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

10 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 sequence selected from the group consisting of SEQ ID NO:1-39, 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:1-39, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:40-78, 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:40-78, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

25 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 selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:40-78, 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:40-78, 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 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

WO 01/42288

PCT/US00/33382

optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence
5 selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:40-78, 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:40-78, 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
10 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
15 selected from the group consisting of SEQ ID NO:1-39, 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:1-39, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and a pharmaceutically acceptable excipient. In one
20 embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-39. 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
25 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:1-39, 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:1-39, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and d) an immunogenic fragment of an amino
30 acid sequence selected from the group consisting of SEQ ID NO:1-39. 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
35 functional GCREC, comprising administering to a patient in need of such treatment the composition.

WO 01/42288

PCT/US00/33382

Additionally, the invention provides a method for screening a compound for effectiveness as 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:1-39, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence
5 selected from the group consisting of SEQ ID NO:1-39, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39. 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
10 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 GCRC, 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
15 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:1-39, 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:1-39, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and d) an immunogenic fragment of an amino acid
20 sequence selected from the group consisting of SEQ ID NO:1-39. 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
25 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:1-39, 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:1-39, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and d) an immunogenic fragment of an amino
30 acid sequence selected from the group consisting of SEQ ID NO:1-39. 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
35 the polypeptide in the presence of the test compound is indicative of a compound that modulates the

WO 01/42288

PCT/US00/33382

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:40-78, 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:40-78, 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:40-78, 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:40-78, 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:40-78, 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 FIGURES AND TABLES

Figure 1 shows the hydrophobicity plot for GCREC-1 (SEQ ID NO:1; Incyte ID number 104941CD1). The hydrophobicity plot was generated using the MacDNASIS Pro software. The positive X axis reflects amino acid position, and the negative Y axis, hydrophobicity. The numbers indicate the positions of predicted transmembrane domains.

Figure 2 shows the hydrophobicity plot for GCREC-3 (SEQ ID NO:3; Incyte ID number 3168839CD1).

WO 01/42288

PCT/US00/33382

Figure 3 shows the hydrophobicity plot for GCREC-4 (SEQ ID NO:4; Incyte ID number 3291235CD1).

Figure 4 shows the hydrophobicity plot for GCREC-5 (SEQ ID NO:5; Incyte ID number 7472001CD1).

5 Figure 5 shows the hydrophobicity plot for GCREC-6 (SEQ ID NO:6; Incyte ID number 7472003CD1).

Figure 6 shows the hydrophobicity plot for GCREC-7 (SEQ ID NO:7; Incyte ID number 7472004CD1).

10 Figure 7 shows the hydrophobicity plot for GCREC-19 (SEQ ID NO:19; Incyte ID number 3068234CD1).

Figure 8 shows the hydrophobicity plot for GCREC-20 (SEQ ID NO:20; Incyte ID number 5029478CD1).

Figure 9 shows the hydrophobicity plot for GCREC-21 (SEQ ID NO:21; Incyte ID number 5102576CD1).

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

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

20 Table 3 shows structural features of each polypeptide sequence, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of each polypeptide.

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

25 Table 5 shows the representative cDNA library for each polynucleotide of the invention.

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

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

30

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing

WO 01/42288

PCT/US00/33382

particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

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

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

DEFINITIONS

"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.

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.

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. 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 polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of

WO 01/42288

PCT/US00/33382

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.

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

- The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

- "Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

- The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of 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 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 fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

- The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that

WO 01/42288

PCT/US00/33382

makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic 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,

WO 01/42288

PCT/US00/33382

Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to
 5 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
 10 conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Glu, His
15	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
20	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
25	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
30	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation,
 (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the
 35 side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl,
 40 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

WO 01/42288

PCT/US00/33382

by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

5 A "fragment" is a unique portion of GCREC or the polynucleotide encoding GCREC 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
10 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
15 in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:40-78 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:40-78, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:40-78 is useful, for
20 example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:40-78 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:40-78 and the region of SEQ ID NO:40-78 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:1-39 is encoded by a fragment of SEQ ID NO:40-78. A fragment
25 of SEQ ID NO:1-39 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-39. For example, a fragment of SEQ ID NO:1-39 is useful as an immunogenic peptide _____

WO 01/42288

PCT/US00/33382

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

5 achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191.

10 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 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

20 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/ncbi2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST

25 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

30

WO 01/42288

PCT/US00/33382

Filter: on

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

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

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

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

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

WO 01/42288

PCT/US00/33382

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
5 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
10 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
15 complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e.,
20 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)
25 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
30 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

WO 01/42288

PCT/US00/33382

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., C₃ or R₃ analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of 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,

WO 01/42288

PCT/US00/33382

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 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*

WO 01/42288

PCT/US00/33382

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).

5 Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South
10 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
15 binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge
20 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
25 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
30 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.

WO 01/42288

PCT/US00/33382

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.

5 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.

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

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

The term "sample" is used in its broadest sense. A sample suspected of containing 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.

20 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 25 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 30 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,

WO 01/42288

PCT/US00/33382

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.

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

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

20 A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may

WO 01/42288

PCT/US00/33382

possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

15 THE INVENTION

The invention is based on the discovery of new human G-protein coupled receptors (GPCRE), the polynucleotides encoding GPCRE, 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.

20 Table 1 summarizes the nomenclature for the polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

25 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 each polypeptide of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. 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.

35 Table 3 shows various structural features of each of the polypeptides of the invention. Columns

WO 01/42288

PCT/US00/33382

1 and 2 show the polypeptide sequence identification number (SEQ ID NO.) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

As shown in Table 4, the 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.40-78 or that distinguish between SEQ ID NO.40-78 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 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 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, 927003T6 is the identification number of an Incyte cDNA sequence, and BRAINOT04 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., 70489898V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g835247) which contributed to the assembly of the polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, g4190944.v113.gs_10.edit is the identification number of a Genscan-predicted coding sequence, with g4190944 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. (See Example V.) Alternatively, the identification numbers in column 5

WO 01/42288

PCT/US00/33382

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.

5 Table 5 shows the representative cDNA libraries for those 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.

10 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
15 embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:40-78, which encodes GCREC. The polynucleotide sequences of SEQ ID NO:40-78, 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.

20 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:40-78 which has at
25 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:40-78. 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.

30 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

WO 01/42288

PCT/US00/33382

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:40-78 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols

WO 01/42288

PCT/US00/33382

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 GCREC 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 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

WO 01/42288

PCT/US00/33382

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 MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) *Nat. Biotechnol.* 17:793-797; Christians, F.C. et al. (1999) *Nat. 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 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., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser.* 7:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser.* 7:225-232.) Alternatively, GCREC itself or a fragment thereof may be synthesized using chemical methods. For example, peptide

WO 01/42288

PCT/US00/33382

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 (1996) 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 enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

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

WO 01/42288

PCT/US00/33382

A variety of expression vector/host systems may be utilized to contain and express sequences encoding GCRC. 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);

5 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*

10 *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

15 Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding GCRC. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding GCRC can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding GCRC into the vector's multiple cloning site

20 disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed

25 bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of GCRC are needed, e.g. for the production of antibodies, vectors which direct high level expression of GCRC may be used. For example, vectors containing

30 the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of GCRC. 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

35 of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, *supra*;

WO 01/42288

PCT/US00/33382

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 alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding GCREC may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses 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 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 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 being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase

WO 01/42288

PCT/US00/33382

genes, for use in *rK*⁺ and *apr*^r 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

5 chlorosulfuron 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

10 glucuronidase and its substrate β -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is

15 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

20 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

25 amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of 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

30 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., Hampton, R. et al. (1990) *Serological Methods, a Laboratory Manual*, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) *Current Protocols in Immunology*, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) *Immunochemical Protocols*, Humana Press, Totowa NJ.)

WO 01/42288

PCT/US00/33382

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.

- 5 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
10 (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
15 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
20 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
25 (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

- In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding GCREC may be ligated to a heterologous sequence resulting in translation of a
30 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

WO 01/42288

PCT/US00/33382

binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion 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 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

WO 01/42288

PCT/US00/33382

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
5 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
10 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.

15 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
20 and grown in culture. The ES cells are transformed with a vector containing the gene of interest
disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989)
Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by
homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP
system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D.
25 (1996) *Clin. Invest.* 97:1999-2002; Wagner, K.U. et al. (1997) *Nucleic Acids Res.* 25:4323-4330).
Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from
the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the
resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains.
Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

30 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
into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998)
Science 282:1145-1147).

35 Polynucleotides encoding GCREC can also be used to create "knockin" humanized animals

WO 01/42288

PCT/US00/33382

(pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding GCRC is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress GCRC, e.g., by secreting GCRC in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of GCRC and G-protein coupled receptors. In addition, the expression of GCRC is closely associated with cancerous, neurological, gastrointestinal, and lung tissue. Therefore, GCRC appears to play a role in cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections. In the treatment of disorders associated with increased GCRC expression or activity, it is desirable to decrease the expression or activity of GCRC. In the treatment of disorders associated with decreased GCRC expression or activity, it is desirable to increase the expression or activity of GCRC.

Therefore, in one embodiment, GCRC 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 GCRC. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Strausler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental

WO 01/42288

PCT/US00/33382

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

5 gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease,

10 aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebotrombosis, vascular 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,

15 rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic 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

20 carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal 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,

25 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,

30 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

35 hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal

WO 01/42288

PCT/US00/33382

dystrophy (APECED), bronchitis, cholecystitis, contact 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, 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.

15 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.

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.

25 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.

35 In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary

WO 01/42288

PCT/US00/33382

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 lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guérin) and *Corynebacterium parvum* are especially preferable.

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.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate

WO 01/42288

PCT/US00/33382

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 idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for 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 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*).

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 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

WO 01/42288

PCT/US00/33382

ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification 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).

5 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 antibody quality and usage in various applications, are generally available. (See, e.g., Catty, *supra*, and Coligan et al. *supra*.)

10 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, 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.)

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

20 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

WO 01/42288

PCT/US00/33382

immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaise, 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. Soma (1997) *Nature* 389:239-242), (i) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (ii) 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 GCRC expression or regulation causes disease, the expression of GCRC 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 GCRC are treated by constructing mammalian expression vectors encoding GCRC and introducing these vectors by mechanical means into GCRC-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R. A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J.-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of GCRC 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). GCRC may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) *Proc. Natl. Acad. Sci. U.S.A.* 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 ocdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V.

WO 01/42288

PCT/US00/33382

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
5 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.

10 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
15 required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. U.S.A.* 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.
20 (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.
25 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. U.S.A.* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

30 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

WO 01/42288

PCT/US00/33382

(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 Antinuzzi, P.A. et al. (1999) *Ann. Rev. Nutr.* 19:511-544 and Verma, I.M. and N. Somia (1997) *Nature* 18:389:239-242, both incorporated by reference herein.

5 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.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV δ 92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

25 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

WO 01/42288

PCT/US00/33382

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
5 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
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., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr,
15 Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme
20 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.

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,
25 GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

30 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 encoding GCREC. Such DNA sequences may be incorporated into a wide variety of vectors with

WO 01/42288

PCT/US00/33382

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 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, 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 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 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

WO 01/42288

PCT/US00/33382

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

5 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 HsLa 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

10 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

15 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

20 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. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various

25 formulations are commonly known and are thoroughly discussed in the latest edition of *Remington's Pharmaceutical Sciences* (Mack Publishing, Easton PA). Such compositions may consist of GCREC, antibodies to GCREC, and nematics, agonists, antagonists, or inhibitors of GCREC.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal,

30 intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, 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 of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting

35 formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins),

WO 01/42288

PCT/US00/33382

recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

5 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 macromolecules comprising GCREC or fragments thereof. For example, liposome preparations
10 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 al. (1999) Science 285:1569-1572).

15 For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

20 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
25 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.
30 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

WO 01/42288

PCT/US00/33382

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.

5 Normal dosage amounts may vary from about 0.1 μg to 100,000 μg , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells,
10 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
15 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 wide variety of reporter molecules, several of which are described above, are known in the art and may
20 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
25 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
30 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.

WO 01/42288

PCT/US00/33382

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:40-78 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 ³²P or ³⁵S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding 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 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 disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Strausler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the

WO 01/42288

PCT/US00/33382

nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental 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,

5 peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial

10 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 angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease,

15 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 erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; a gastrointestinal

20 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, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis,

25 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, alpha-

30 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

35 disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia,

WO 01/42288

PCT/US00/33382

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, hyper eosinophilia, 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, hepatitis virus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and togavirus. The polynucleotide sequences encoding GCREC 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 fluids or tissues from patients to detect altered GCREC expression. Such qualitative or quantitative methods are well known in the art.

20 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.

30 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

WO 01/42288

PCT/US00/33382

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,

- 5 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.

- 10 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.

- 15 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 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.

- 20 Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

- In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding 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 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 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

WO 01/42288

PCT/US00/33382

overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

5 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 standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplax, 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

10 quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large

15 genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used

20 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

25 may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by

30 quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the

WO 01/42288

PCT/US00/33382

hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

5 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.

10 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
15 molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaisir, E.F. et al. (1999) *Mol. Carcinog.* 24:153-159; Steiner, S. and N.L. Anderson (2000) *Toxicol. Lett.* 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties.
20 These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data
25 after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is
important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated
30 biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present
35 invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global

WO 01/42288

PCT/US00/33382

pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and
5 analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are
10 visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for
15 example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for GCREC to quantify the
20 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; Mendez, 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-
25 reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the
30 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

WO 01/42288

PCT/US00/33382

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., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shaloo, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in *DNA Microarrays: A Practical Approach*, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding GCRC may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM)

WO 01/42288

PCT/US00/33382

World Wide Web site. Correlation between the location of the gene encoding GCRC 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., Oatis, 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, GCRC, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between GCRC and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with GCRC, or fragments thereof, and washed. Bound GCRC is then detected by methods well known in the art. Purified GCRC can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding GCRC specifically compete with a test compound for binding GCRC. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with GCRC.

In additional embodiments, the nucleotide sequences which encode GCRC 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

WO 01/42288

PCT/US00/33382

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
5 in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, in particular U.S. Ser. No. 60/172,852, U.S. Ser. No. 60/171,732, U.S. Ser. No. 60/176,148, and U.S. Ser. No. 60/177,331, are expressly incorporated by reference herein.

10

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. The Incyte cDNA shown for SEQ ID NO:40 was derived from a cDNA library constructed from bone marrow tissue. The Incyte cDNAs
15 shown for SEQ ID NO:41 were derived from cDNA libraries constructed from small intestine, including tissues associated with Crohn's disease, from large intestine, and from brain tissues. The Incyte cDNAs shown for SEQ ID NO:42 were derived from cDNA libraries constructed from prostate tumor, small intestine, breast, and epidermal tissues. The Incyte cDNAs shown for SEQ ID NO:43 were derived cDNA libraries constructed from soft tissue tumor, fetal rib, and brain tissue associated
20 with Huntington's disease. The Incyte cDNAs shown for SEQ ID NO:57 were derived from cDNA libraries constructed from lymphocytes and mast cells, and from breast, uterine, prostate, adrenal gland, spinal cord, tibial muscle, lung, esophagus, small intestine, and colon tissues. The Incyte cDNAs shown for SEQ ID NO:58 were derived from cDNA libraries constructed from a fallopian tube tumor, uterine endometrium, and bronchial tissue. The Incyte cDNAs shown for SEQ ID NO:59 were derived
25 from cDNA libraries constructed from colon tissues, including cecal tumor tissue, as well as from pancreatic tumor, pituitary gland, and brain tissues. The Incyte cDNAs shown for SEQ ID NO:60 were derived from cDNA libraries constructed from brain, including brain tumor tissue and tissues associated with Huntington's disease, and from prostate tumor, cervical adenocarcinoma, breast, small intestine, and bladder tissues. Some tissues were homogenized and lysed in guanidinium isothiocyanate,
30 while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.
35 Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA

WO 01/42288

PCT/US00/33382

purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSOFT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*, units 5.1-5.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESOFT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), or pINCY (Incyte Genomics, Palo Alto CA). Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5a, DH10B, or ElectroMAX DH10B from Life Technologies.

20 II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) *Anal. Biochem.* 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

WO 01/42288

PCT/US00/33382

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

5 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

10 carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

15 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,

20 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 HMMR. The Incyte cDNA sequences were assembled to produce polynucleotide sequences. Alternatively, GenBank cDNAs,

25 GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Pfrod, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The polynucleotide sequences were translated to derive the corresponding polypeptide sequences which were

30 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

35 generated using default parameters specified by the CLUSTAL algorithm as incorporated into the

WO 01/42288

PCT/US00/33382

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 assembled polynucleotide 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 polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:40-78. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

15 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., gbpr1 and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) *J. Mol. Biol.* 268:78-94, and Burge, C. and S. Karlin (1998) *Curr. Opin. Struct. Biol.* 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode G-protein coupled receptors, the encoded polypeptides were analyzed by querying against PFAM models for G-protein coupled receptors (7tm_1, 7tm_2, 7tm_3, and 7tm_4). Potential G-protein coupled receptors were also identified by homology to Incyte cDNA sequences that had been annotated as G-protein coupled receptors. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpr1 public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Polynucleotide sequences, including SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:61, SEQ ID NO:62,

WO 01/42288

PCT/US00/33382

and SEQ ID NO:63, 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, polynucleotide sequences, including SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, 5 SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, and SEQ ID NO:78, are full length coding regions derived entirely from edited or unedited Genscan-predicted coding sequences. Alternatively, polynucleotide sequences, including SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, and SEQ ID NO:56, are partial genes 10 derived from the assembly and editing of Genscan-predicted sequences only.

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 15 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 20 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 25 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 30 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 01/42288

PCT/US00/33382

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 GCRC Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO.40-78 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.40-78 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

Map locations are represented by ranges, or intervals, 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 Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

WO 01/42288

PCT/US00/33382

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar.

5 The basis of the search is the product score, which is defined as:

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

10 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
 15 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
 20 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 GCREC 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
 25 derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of
 30 libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA
 35 encoding GCREC. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ

WO 01/42288

PCT/US00/33382

GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of GCRC 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, 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

WO 01/42288

PCT/US00/33382

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

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

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

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:40-78 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 µCi of [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 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 membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and

WO 01/42288

PCT/US00/33382

compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing. See, e.g., Baldeschweiler, supra), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schna (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., Schna, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

25 Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 μ l volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified

WO 01/42288

PCT/US00/33382

using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and

5 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

10 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

15 washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

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

20 Patent No. 5,807,522, incorporated herein by reference. 1 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.

25 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

30 Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for

35 about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC,

WO 01/42288

PCT/US00/33382

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 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 typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, 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 measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping 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).

WO 01/42288

PCT/US00/33382

XI. Complementary Polynucleotides

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

10 XII. Expression of GCREC

Expression and purification of GCREC is achieved using bacterial or virus-based expression systems. For expression of GCREC in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express GCREC upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of GCREC 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 GCREC 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, GCREC is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from GCREC at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification

WO 01/42288

PCT/US00/33382

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

5 Examples XVI, XVII, and XVIII, 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

10 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

15 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

20 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;

25 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

30 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

WO 01/42288

PCT/US00/33382

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-iodinated 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

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 three extracellular loops, the extracellular N-terminal region, or the third intracellular loop. Candidate

WO 01/42288

PCT/US00/33382

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.

5 Alternatively, molecules interacting with GCREC are analyzed using the yeast two-hybrid system as described in Fields, S. and D. 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
10 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
15 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-
20 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
25 (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 nGEX (Pharmacia Biotech). The construct is transformed into

WO 01/42288

PCT/US00/33382

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. These

5 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% NaN₃, 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

10 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).

15 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

20 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

25 [³H]thymidine into acid-precipitable DNA is measured over an appropriate time interval using a radioisotope 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 concentration of GCREC producing a 50% response level, where 100% represents maximal

30 incorporation of [³H]thymidine into acid-precipitable DNA (McKay, J. 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 GCREC is

35 transfected into a mammalian cell line (e.g., Chinese hamster ovary (CHO) or human embryonic

WO 01/42288

PCT/US00/33382

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 GCRC present in the transfected cells.

To measure changes in inositol phosphate levels, the cells are grown in 24-well plates containing 1×10^6 cells/well and incubated with inositol-free media and [3 H]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 GCRC present in the transfected cells.

XVIII. Identification of GCRC Ligands

GCRC 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 GCRC 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 Ca^{2+} . 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 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, GCRC may be coexpressed with the G-proteins $G_{\alpha 13/14}$ which have been demonstrated to couple to a wide range of G-proteins (Offermanns, S. and M.I. Simon (1995) J. Biol. Chem. 270:15175-15180), in order to funnel the signal transduction of the GCRC through a pathway involving phospholipase C and Ca^{2+} mobilization. Alternatively, GCRC may be expressed in engineered yeast systems which lack endogenous GPCRs, thus providing the advantage of a null background for GCRC activation screening. These yeast systems substitute a

WO 01/42288

PCT/US00/33382

human GPCR and Gα 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. Thorner (1996) Nature 384 (supp.):14-16). The receptors are screened against putative
5 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.

10 Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

WO 01/42288

PC/T/US00/33382

Table 1

Project ID	incyte	Polypeptide SEQ ID NO:	incyte Polypeptide ID	incyte Polynucleotide SEQ ID NO:	incyte Polynucleotide ID
1499408		1	104941CD1	40	104941CB1
3168339		2	1499408CD1	41	1499408CB1
3291235		3	3168339CD1	42	3168339CB1
7472001		4	3291235CD1	43	3291235CB1
7472003		5	7472001CD1	44	7472001CB1
7472004		6	7472003CD1	45	7472003CB1
7472005		7	7472004CD1	46	7472004CB1
7472006		8	7472005CD1	47	7472005CB1
7472007		9	7472006CD1	48	7472006CB1
7472008		10	7472007CD1	49	7472007CB1
7472009		11	7472008CD1	50	7472008CB1
7472010		12	7472009CD1	51	7472009CB1
7472011		13	7472010CD1	52	7472010CB1
7472012		14	7472011CD1	53	7472011CB1
7472013		15	7472012CD1	54	7472012CB1
7472014		16	7472013CD1	55	7472013CB1
7472015		17	7472014CD1	56	7472014CB1
7472016		18	7472015CD1	57	7472015CB1
7472017		19	7472016CD1	58	7472016CB1
7472018		20	7472017CD1	59	7472017CB1
7472019		21	7472018CD1	60	7472018CB1
7472020		22	7472019CD1	61	7472019CB1
7472021		23	7472020CD1	62	7472020CB1
7472022		24	7472021CD1	63	7472021CB1
7472023		25	7472022CD1	64	7472022CB1
7472024		26	7472023CD1	65	7472023CB1
7472025		27	7472024CD1	66	7472024CB1
7472026		28	7472025CD1	67	7472025CB1
7472027		29	7472026CD1	68	7472026CB1

WO 01/42288

PCT/US00/33382

Table 1 (cont.)

Project ID	Incyte	Polymerptide SEQ ID NO:	Incyte	Polymerptide ID	Polymerptide SEQ ID NO:	Incyte
7472017		30	7472017CDB		69	7472017CBI
7472018		31	7472018CDB		70	7472018CBI
7472019		32	7472019CDB		71	7472019CBI
7472021		33	7472021CDB		72	7472021CBI
7472029		34	7472029CDB		73	7472029CBI
7472010		35	7472010CDB		74	7472010CBI
7472011		36	7472011CDB		75	7472011CBI
7472014		37	7472014CDB		76	7472014CBI
7472015		38	7472015CDB		77	7472015CBI
7472020		39	7472020CDB		78	7472020CBI

WO 01/42288

PC/T/US00/33382

Table 2

Polypeptide SEQ ID NO.	Incyte Polypeptide ID	Genbank ID NO.	Probability Score	Genbank Homolog
1	104941CD1	57211116	6.9E-146	Olfactory receptor [Callithrix jacchus]
2	349408CD1	5202806	5.20E-162	Vasopressin receptor [Rattus norvegicus]
3	316839CD1	53618229	2.2E-44	G protein-linked P2Y4 receptor [Rattus norvegicus]
4	229123CD1	53287459	1.40E-148	A-2 [Mus musculus]
5	247030CD1	5157594	4.20E-136	Taste bud receptor protein TB 641 [Rattus norvegicus]
6	247204CD1	5157594	4.20E-136	G protein-coupled receptor-like protein (GPCR) [Rattus norvegicus]
7	747204CD1	5169892	6.30E-138	High affinity lysophosphatidic acid receptor [Rattus norvegicus]
8	7475887CP1	51256393	4.70E-90	Taste bud receptor protein TB 641 [Rattus norvegicus]
9	7483028CP1	52447219	2.50E-75	OLF4 [Homo sapiens]
10	747933CP1	52792016	4.90E-75	Olfactory receptor [Homo sapiens]
11	7475164CP1	5517366	3.00E-111	Olfactory receptor [Rattus norvegicus]
12	7473209CP1	54680264	1.50E-32	Olfactory receptor 525 [Mus musculus]
13	747525CP1	52447219	6.50E-61	OLF1 [Homo sapiens]
14	792757CP1	5810089	6.7E-54	Putative taste receptor TR2 [Homo sapiens]
15	7481257CP1	54826521	4.00E-29	OR10B1 (novel 7 transmembrane receptor (rhodopsin family)) [Homo sapiens]
16	7485790CP1	52447219	3.00E-40	Olfactory receptor like protein (h56H1-15) [Homo sapiens]
17	7487993CP1	51114665	3.10E-34	OLF2 [Homo sapiens]
18	2829053CD1	51114665	3.10E-34	OLF3 [Canis familiaris]
19	3068234CD1	55922725	3.1E-130	Lysophosphatidic acid G protein-coupled receptor (Homo sapiens)
20	5029478CD1	51049072	3.6E-21	Galanin receptor GALR1 (Rattus norvegicus) (Cloning and characterization of the rat GALR1 galanin receptor from Rln149 insulinoma cells. Brain Res. Mol. Brain Res. 1995 Dec 28;34(2):179-189.)

WO 01/42288

PC/T/US00/33382

Table 2 (cont.)

POLYPEPTIDE SEQ ID NO.	Incyte Polypeptide ID	GenBank ID NO.	Probability Score	GenBank Homolog
21	5102576CD1	32792016	2.4E-92	Olfactory receptor (Homo sapiens) (Molecular cloning and chromosomal mapping of olfactory receptor genes expressed in the male germ line: evidence for their wide distribution in the mouse genome). Proc Natl Acad Sci U S A. 1997; 94(18):23703-23708.
22	32200334CD1	35003404	4.6E-131	57kD, 15 (tm17M1-6) 7-transmembrane olfactory receptor-like protein (ichthodopsin family) [Mus musculus]
23	3275821CD1	4182742	1.5E-29	Formyl peptide receptor (Homo sapiens) [Murphy, P.M. et al. (1992) J. Biol. Chem. 267:7637-7643]
24	3744167CD1	59186902	1.2E-240	Leukotriene B4 receptor, BLT2 [Mus musculus]
25	7472007CD1	57538409	1.3E-159	Olfactory receptor P2 [Mus musculus]
26	7472008CD1	94218182	1.0E-89	OR71M21.2 (hs6M-12) 7 transmembrane receptor (rhodopsin family) (olfactory receptor like) (orotelin) [Homo sapiens]
27	7472013CD1	4205846	2.5E-70	Olfactory protein [Rattus norvegicus]
28	7472014CD1	5670948	2.5E-73	Olfactory receptor (Gorilla gorilla)
29	7472016CD1	56090736	1.1E-215	Olfactory receptor (Gorilla gorilla)
30	7472017CD1	33757727	2.0E-61	OR6M19.7 (olfactory receptor-like protein (hs6M1-3)) [Homo sapiens]
31	7472018CD1	56544328	2.3E-112	Orphan G protein-coupled receptor GPR26 [Rattus norvegicus]
32	7472019CD1	55859916	2.7E-73	Olfactory receptor [Mus musculus]
33	7472021CD1	56090804	2.6E-94	Olfactory receptor (Gorilla gorilla)
34	7472009CD1	51016362	1.6E-68	OLI receptor [Rattus norvegicus]
35	7472010CD1	52317704	7.3E-80	Olfactory receptor [Rattus norvegicus]
36	7472011CD1	56178008	4.9E-114	Olfactory receptor MOR18 [Mus musculus]
37	7472012CD1	4205816	6.8E-84	Olfactory protein [Rattus norvegicus]
38	7472014CD1	4205816	3.9E-88	Olfactory protein [Rattus norvegicus]
39	7472020CD1	52792016	1.9E-97	Olfactory receptor (Homo sapiens)

WO 01/42288

PC/T/US00/33382

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	104941CD1	311	S68 S189 S292 Y310	N5 N66	7 transmembrane receptor (rhodopsin family); G42-L291 GPCR signature; K81-F130, L208-Y219, Y283-K299, Y103-S151 Olfactory receptor signature; M60-K81, F178-D192, F219-G254, A275-L286, S292-Q306 Transmembrane domains: J31-L47, F241-L229	MOTIFS HMMER-PFAM BLOCKS BLIMPS PRINTS ProfileScan
2	149940CD1	891	T146 S266 S114 S48 S56 S61 S401 T628 T814 S856 T84 S146 T144 T325 T411 T541 S568 S676 T706 T789 Y372	N178	MP7/GTP binding site (P-loop); G202-T209	MOTIFS
3	318883CD1	422	T332 S178 T342 S163 S171 S397 T311 S126 T307 S332 S367	N4 N9 N251 N323	7 transmembrane receptor (rhodopsin family); L93-L297 Rhodopsin-like GPCR superfamily; R14-L48, V57-R78, F101-I123, V137-R158, V192-F215, T232-V256, L279-R305	MOTIFS HMMER-PFAM BLOCKS BLIMPS ProfileScan PRINTS HMMER
4	325123CD1	609	S228 S229 S356 S456 S724 S328 S164 S417 S466 S706 S569 S390 S292 S315 S462 S489 S560 Y348		7 transmembrane receptor (rhodopsin family); E90-E154 GPCR signature; R14-L48, V57-R78, F101-I123, V137-R158, V192-F215, T232-V256, L279-R305 Transmembrane domains: V275-L295	MOTIFS HMMER-PFAM BLOCKS BLIMPS ProfileScan PRINTS HMMER

WO 01/42288

PC/T/US00/33382

Table 3 (cont.)

SRC ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	7472004CD1	313	S88 T194 T200 S247 T309 T138 T164 T290 S306	N5 N85	7 transmembrane receptor (rhodopsin family): G41-I239 GPCR signature: K81-E130, T281-K297, Y103-R147 Olfactory receptor signature: M60-R61, F178-D192, F239-V254, A273-L284, T290-L304 Signal peptide: M1-T38	HMMER-PFAM MOTIFS BLIMPS- BLIMPS- PRINTS ProfileScan HMMER SPScan
6	7472003CD1	398	S16 T155	N10 N18 N28	Transmembrane domains: F30-T49, F63-M63 7 transmembrane receptor (rhodopsin family): G45-V233 GPCR signature: W108-R147, Y213-Y224, A256-F282, N315-R331, M119-I166 Neuropeptide Y receptor signature: R69-T81, L321-F334	MOTIFS HMMER-PFAM BLIMPS- BLIMPS- PRINTS ProfileScan
7	7472004CD1	369	S228 T34 T218 S159 T150	M2	Transmembrane domain: A2-Y45 7 transmembrane receptor (rhodopsin family): G46-Y321 Rhodopsin-like GPCR signature: T13-V57, I66-F67, F111-I133, R144-V165, V191-L216, A242-V286, S103-R329 Transmembrane domains: T11-V51, M109-I125, Y189-M213, M256-V275	HMMER MOTIFS HMMER-PFAM BLIMPS- BLIMPS- PRINTS HMMER
8	7475667CF1	194	T186 T76 T62 T46 T172		7 transmembrane receptor (rhodopsin family): M1-Y171 Opsins retinal binding site: Y142-M194 Olfactory receptor signature: F60-D74, F121-V136, A155-L166, T172-T186	MOTIFS HMMER-PFAM BLIMPS- BLIMPS- PRINTS ProfileScan

WO 01/42288

PC/T/US00/33382

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
9	7483029CPI	173	T16 S34 T60	N32 N187	7 transmembrane receptor (rhodopsin family) L146 Rhodopsin-like GPCR signature: M26-K47, F71-193, L107-I128 Signal peptide: M1-L22	NOTIFS HMMER-PFAM BLOCKS ProfileScan BLIMPS-PRINTS SPScan HMMER
10	7477933CPI	220	S172		7 transmembrane domains: 7 transmembrane receptor (rhodopsin family) p1-C192 Olfactory receptor signature: M2-K23, F120-D134, F181-G196 Rhodopsin-like GPCR signature: L47-169, I142-V165	NOTIFS HMMER-PFAM BLOCKS BLIMPS-PRINTS HMMER
11	7475164CPI	302	T395 S58 S84 T107 T257 T9 T69 S128 T151 S282		7 transmembrane receptor (rhodopsin family) G12-I193 GPCR signature: N81-F120, I273-K289, S93-L142 Olfactory receptor signature: V50-K71, Y168-S182, F229-G344, S255-I276, S282-V296 Transmembrane domains: M41-R62, V85-T110	NOTIFS HMMER-PFAM BLOCKS ProfileScan BLIMPS-PRINTS HMMER
12	7475939CPI	110	S70 S38 T66 S94		GPCR signature: I85-K101 Olfactory receptor signature: F41-G56, A77-L88, S94-Y108	NOTIFS BLOCKS BLIMPS-PRINTS HMMER

WO 01/42288

PC/T/US00/33382

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
13	7475252CF1	178	S66 S151 S136	N4 N64	7 transmembrane receptor (rhodopsin family): G40-L153 Rhodopsin-like GPCR signature: V95-S49, K58-K79, L103-I129, S102-S151	HMER-PFAM MOTIFS BLIMPS- BLOCKS ProfileScan BLIMPS- PRINTS HMER
14	7927572CF1	92			Transmembrane domains: L29-I45, M100-M117 Olfactory receptor signature: F25-D39	HMER MOTIFS BLIMPS- PRINTS
15	7481257CF1	97			7 transmembrane receptor (rhodopsin family): M1-Y96 GPCR signature: V13-Y24, D41-Q67 Olfactory receptor signature: F44-G59, I80-I91 Signal peptide: M1-G27	MOTIFS HMER-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS SFSscan HMER
16	7485790CF1	133	S74		Transmembrane domain: M8-Y24 Transmembrane receptor (rhodopsin family): R22-V128 GPCR signature: G71-F110	MOTIFS HMER-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS HMER
17	7482992CF1	213	S85 S205 S159 T183	N63	Transmembrane domain: M82-A100 7 transmembrane receptor (rhodopsin family): S2-V182 GPCR signature: R127-S153, S2-K38 Olfactory receptor signature: F69-N63, F130-G145, V166-L177, T183-G197 Transmembrane domains: F102-I120, F130-V152	MOTIFS HMER-PFAM BLIMPS- BLOCKS ProfileScan BLIMPS- PRINTS HMER

WO 01/42288

PCT/US00/33382

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
18	2829053CD1	180	S30 S41 S109 S125 S140 S35 S36 S145		Beta-1 adrenergic receptor signature: T146-S146 7 transmembrane domain: M1-S57 7 transmembrane receptor (rhodopsin family): S47-Y293	MOTIFS BLIMPS- PRINTS
19	3068234CD1	353	T146 T117 T233 S221 S17 T21 S294 S293 T141 S229 T303 Y14	M15 M139 M172 N249	Rhodopsin-like GPCR superfamily signature: I22-I56, F65-L96, L109-I131, R144-L165, Y187-Y210, L217-L261, K228-I301 7 transmembrane domains: Y36-T56, T146-S146, Y187-L207, T240-V258	MOTIFS HMMER-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS
20	5029478CD1	361	T242 S256 S237 S350	M21 N322	7 transmembrane receptor (rhodopsin family): G57-Y221 Rhodopsin-like GPCR superfamily signature: V42-A66, T74-W95, M18-I140, S148-L165, E208-L231, L202-L286, F303-K329 Transmembrane domains: T45-V45, V124-Q144, V209-I233, L266-K291	MOTIFS HMMER-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS
21	5102576CD1	251	S119 S196		7 transmembrane receptor (rhodopsin family): R8-C251 GPCR signature: Q156 Olfactory receptor signature: M26-K47, L144-D158, F205-G220, A241-G252 Transmembrane domains: T66-D88, V109-T134	MOTIFS HMMER-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS

WO 01/42288

PCT/US00/33382

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
22	2200534CD1	315	S6 S136 T251	N4 N154	G-protein coupled receptor: DM00013(P22750)17-306; S17-L205 DM00013(P22750)17-306; F36-L205 P000009; L68-F169 P0149521; T345-L305; F20-L305 P0000291; C168-L245 P0002495; N4-L47 BL00237; L89-F128, L207-Y218, T282-K288 Difactory receptor PR00245: M84-F19, F179-S190, V238-G253, X274-L265, L245-L205 G-protein coupled receptors	HMMER BLIIPS- BLIIPS- BLOCKS HMMER-PFAM MOTIFS
23	327582LCD1	470	T3 T18 T126 T332 T340 S350 S424 S451 T459 S192	N47	Transmembrane domains: V204-K228, G40-Y290 G-protein coupled receptor: DM00013(P21462)20-317; V14-L306 P000009; L68-F169 BL00237; M7-F136; G201-R212, GFCR profiles: F109-V155 Rhodopsin GPCR family PR00237: W31-G55, L66-S87, W11-A133, L147-V169, L193-Q216, F215-L239, L277-R303 G-protein coupled receptors	BLAST- PRODOM BLAST-DMO HMMER HMMER-PFAM BLIIPS- BLOCKS BLIIPS- MOTIFS ProfileScan
					Transmembrane domains: T33-S51, L68-L259	BLAST- PRODOM BLAST-DMO HMMER

WO 01/42288

PC/T/US00/33382

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
24	3744167CD1	358	T291 S15 T18 S215	N10 N38 N342	G-protein coupled receptor: D80003 (P46021-27-318: 519-F290) D80003 (P46021-27-318: 519-F290) D80003 (P35241-2-323: M16-F290) D80003 (P35241-2-323: M16-F290) E800237: 687-P126, F190-Y201. GPCR profile: R217-V243, S280-L296 Y89-V145 RHO GTPase GPCR family PR00237: T22-A46, A57-F78, C101-V122, L137-V158, L182-L205, V222-L246, R270-L298. G-protein coupled receptors	BLAST- BLIMP2- BLIMP2- BLIMP2- PRF1MS- ProfileScan MOTIFS
					Transmembrane domains: A138-V159, G17-Y288	BLAST- PRODOM BLAST-DOMO HMMER

WO 01/42288

PC/T/US00/33382

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
25	7472007CD1	314	S270 S291 S311 T49 S67 S193	N4 N65	Signal peptide: M1-G56 Transmembrane domains: L23-Y26, L208-R228 7 transmembrane receptor (rhodopsin family) signature: G41-Y290 G-protein coupled receptor signatures: T90-F129, I207-Y213, T282-K299 G-protein coupled receptor signature: Y102-A147 G-protein coupled receptor signature: A110-A125 Olfactory receptor signatures: M59-K80, F177-D191, F238-S253, I274-L285, S29-S105 Melanocortin receptor family: A5-L63 Rhodopsin-like GPCR superfamily: K68-L107, M59-K80, F104-L126, S29-S105, A199-L222, A237-R261, K272-K288 G-protein coupled receptor: DM0013 P23270 L8-311; L23-H306 G-protein coupled receptor: DM0013 P21267 20-309; L27-I105 G-protein coupled receptor: DM0013 P23275 17-306; L23-I105 G-protein coupled receptor: DM0013 P30953 18-308; L49-H306 Olfactory receptor FM009211: F159-L246 Olfactory receptor PD149621: V247-R307	HMER-PPAK R08ER BLIMPS-BLOCKS ProfileScan MOTIFS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-DMO BLAST-DMO BLAST-DMO BLAST-DMO BLAST-DMO BLAST-DMO BLAST-DMO

WO 01/42288

PCT/US00/33382

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
26	747208CD1	365	S78 T192 S199 K226 S243 S247 K262 S286C S217 T222 T337 T346I	N94	Transmembrane domain: I226-L244... 7 transmembrane receptor (rhodopsin family) signature: C170-L199 G-protein coupled receptor signature: K119-R158, L236-S247, K264-Q290, T311-R327 Olfactory receptor signature: M89-Q109, V206-D230, F267-G282, L305-L314, T320-K334 Melanocortin receptor family: K119-L138 Rhodopsin-like GPCR superfamily: L35-L79, M88-Q109, F131-V115, M228-A251, A266-Q290, K301-H327 Olfactory receptor PD000921: L195-L275 Olfactory receptor PD149621: V278-K331 G-protein coupled receptor: DM00013 E30955 18-305; L61-L326 G-protein coupled receptor: DM00013 P23269 15-304; I61-L326 G-protein coupled receptor: DM00013 A57069 15-304; D59-L326 G-protein coupled receptor: DM00013 P23275 17-306; T67-L326	HMER-PPM BLIMPS-BLOCKS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODOM BLAST-PRODOM BLAST-DMO BLAST-DMO BLAST-DMO BLAST-DMO

WO 01/42288

PC/T/US00/33382

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
28	7472015CD1	335	T73 S79 S214 S309 T417 S329 S331		Signal peptide: M1-K20 Signal peptide: M1-K20 Transmembrane domains: T23-V27, V43-V61, R117-I136 Transmembrane domain: T23-V27, V43-V61, R117-I136 Rhodopsin-like GPCR superfamily signatures: T21-Y279 G-protein coupled receptor signatures: R71-F110, F174-Y185, F218-T244, N271-R287 G-protein coupled receptor signature: G24-L28 G-protein coupled receptor signature: A91-I106 Rhodopsin-like GPCR superfamily: S6-L30, S40-L61, V65-I107, V121-G142, V166-L189, A243-V247, E251-E287 G-protein coupled receptor: DM00013 F41398 137-631: G8-D220 G-protein coupled receptor: DM00013 F41398 137-631: G8-D220 G-protein coupled receptor: DM00013 F41800 241-516: G8-Y281 G-protein coupled receptor: DM00013 F21388 20-335: G8-E218 G-protein coupled receptor: DM00013 JRG0591 20-335: G8-E218	HMMER HMMER SPSCAN HMMER-PPAM BLIMPS- BLOCKS ProfileScan MOTIFS BLIMPS- PRIMTS BLAST-DMO BLAST-DMO BLAST-DMO BLAST-DMO

WO 01/42288

PC/T/US00/33382

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
30	7472017CD1	336	37 T217	N5 N189	Signal peptide: MI-342 Transmembrane domains: C11-322, V123-L141 Rhodopsin-like receptor (rhodopsin family) signatures: F12-Y216 G-protein coupled receptor signatures: K24-P63, L133-Y144, C161-Q187, T208-K234 G-protein coupled receptor signatures: R141-L142, T143-L144, T144-A159 G-protein coupled receptor signature: T44-A59 Olfactory receptor signatures: L164-G179, L200-L211, T217-M231 Rhodopsin-like GPCR superfamily: S38-V60, L195-R148, G163-Q187, K188-K224 Olfactory receptor PD149621: T143-L144, T144-A159 Olfactory receptor F000921: C103-L172 G-protein coupled receptor: D400013 P23489 15-304: L15-L227 G-protein coupled receptor: D400013 F30953 18-306: L15-L227 G-protein coupled receptor: D400013 R57069 15-306: L15-R229 G-protein coupled receptor: R330013 F23272 17-306: M1-E227	SWISS-PROT, TrEMBL, HMMER, Pfam, PROSITE, PRODOM, BLAST-DBPROT, BLAST-DBPFI, BLAST-DBMAM, BLAST-DBMOT, BLAST-DBOMO

WO 01/42288

PC/T/US00/33382

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
31	747201UCD1	363	Y294 S325 S326 S327 T240 T240 T316 T340	K47 N348 N355	Signal peptide: M1-A24 transmembrane receptor rhodopsin family signatures: K286-E302 G-protein coupled receptor signatures: W2-P111, F181-S192, R234-T260, K286-E302 Rhodopsin-like GPCR superfamily: L7-A11, S41-F62, D86-V108, Y122-G143, P173-H196, A239-A263, K286-E302 G-protein coupled receptor signatures: P294-E302 G-protein coupled receptor: DMG0013 JN0591 20-336: P2-L305 G-protein coupled receptor: DMG0013 P53452 17-344: L7-F268 G-protein coupled receptor: DMG0013 F50406 20-335: G4-E305 G-protein coupled receptor: DMG0013 P21386 20-336: P2-L305	BLIMPS- BLOCKS BLIMPS- PRINTS BLIMPS- PRINTS BLAST-DMO BLAST-DMO BLAST-DMO BLAST-DMO

WO 01/42288

PC1/US00/33382

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
32	7472019CD1	308	S162 S290 S67 T187 S192 S285	K5 N65	Transmembrane domains: T130-I49, M197-P215 Family signatures: G-protein coupled receptor signatures: P90-P129, L206-Y217, L234-K260, T281-K297 G-protein coupled receptor signature: S102-T147 Olfactory receptor signatures: L59-Q80, F176-Q190, F217-G252, I271-L284, S290-M304 Melanocortin receptor family: F51-L63 Rhodopsin-like GPCR superfamily: T26-H50, I59-Q60, S104-I126, V140-L161, I199-A231, K271-K297 G-protein coupled receptor: DM00013 E23275 I7-306, I17-M304 G-protein coupled receptor: DM00013 E30249 L5-306, F57-M304 G-protein coupled receptor: DM00013 E23266 I7-306, I17-M304 G-protein coupled receptor: DM00013 S23707 L8-306, P21-I300 Olfactory receptor F0000921: L165-L244 Olfactory receptor PD149621: T245-R304	HMMER-PFAM HMMER HMMER-PFAM BLIMPS-BLOCKS BLIMPS-BLOCKS ProfileScan BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-DBO BLAST-DBO BLAST-DBO BLAST-DBO BLAST-PRODOM BLAST-PRODOM

WO 01/42288

PC/T/US00/33382

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
33	7472021CD1	343	S87 T154 S288 S326 S311 S316	N35 N183 N314	Transmembrane domain: Y95-L75, I214-I214 Transmembrane receptor (rhodopsin family) signatures: G61-Y110 G-protein coupled receptor signatures: S38-I64, G310-F149, S302-K318 G-protein coupled receptor signature: F122-V166 G-protein coupled receptor signature: S130-A145 Olfactory receptor signatures: M79-Q100, F297-Y211, F258-S273, F297-L305, S311-L325 Melanocortin receptor family: V71-I83 Vasopressin receptor signature: S175-I86 Olfactory receptor PD000921: S338-S428 Olfactory receptor PD149621: V667-E428 G-protein coupled receptor: DK00013 F23269 I5-304 E40-L325 G-protein coupled receptor: DK00013 F23275 I7-305 S38-L325 G-protein coupled receptor: DK00013 F23273 I8-306 I45-I325 G-protein coupled receptor: DK00013 F23269 I7-306 S38-S326	HMMER-PPA HMMER BLIMPS- BLOCKS ProfileScan MOTIFS BLIMPS- PRINTS BLIMPS- PRINTS BLAST- DOMO BLAST- DOMO BLAST- DOMO BLAST- DOMO

WO 01/42288

PC/T/US00/33382

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
34	7472009C01	323	S87 S232 T290 S8 S67 T193	N5	Transmembrane domains: ? Transmembrane receptor (Rhodopsin family) signatures: G41-V289 G-protein coupled receptor signatures: K90-F129, L206-Y217, R234-R260, T281-A297 G-protein coupled receptor signature: F102-R147 G-protein coupled receptor signature: R440-Q442, A110-A125 Olfactory receptor signatures: M59-K80, F177-D191, F237-G252, G273-L284, T290-L304 Rhodopsin-like GPCR superfamily: F26-C50, Y104-I126, V140-A161, T198-L221, K271-K297 Metanodopsin receptor family: R284-K306, R308-K330 Vasopressin receptor signature: L55-L66 G-protein coupled receptor: DM00013 P23275 17-306; L25-L304 G-protein coupled receptor: DM00013 A57069 15-304; L27-L304 G-protein coupled receptor: DM00013 P23270 18-311; L25-L304 G-protein coupled receptor: O1F6C0033 P23269 17-306; L27-L304 Olfactory receptor F0419621: T245-T310 Olfactory receptor F0000921: F168-L244	HMWER-PFAM HMWER BLIMPS-BLOCKS ProfileScan MOTIFS BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-DMO BLAST-DMO BLAST-DMO BLAST-DMO BLAST-PRODOM PRODOM

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
35	7472H10CD1	299	T68, S128, S280, T293, S10, S57, I136	N95	Transmembrane domain: 7 L186-2205 Transmembrane receptor (Rhodopsin family) signatures: G31-Y279 G-protein coupled receptor signatures: S79-F118, F188-S199, S224-T250, Y271-S287 G-protein coupled receptor signature: F91-F135 G-protein coupled receptor signature: S199-S287 S89-S114 Olfactory receptor signatures: M49-K70, Y166-S180, F227-G242, A163-L274, S280-L294 Melanocortin receptor family: I91-I153 Rhodopsin-like GPCR superfamily: I490-S185 M49-K70, Y166-S180, F227-G242, M181-L264, S267-T250, F262-K287 G-protein coupled receptor: DM00013 S29709 11-299: G23-L294 G-protein coupled receptor: DM00013 S51356 18-307: I24-K292 G-protein coupled receptor: DM00013 P23274 19-305: I24-L294 G-protein coupled receptor: DM00013 P30955 18-305: I24-L294 Olfactory receptor PD149621: V195-S286 Olfactory receptor PD000921: L159-I235	MEMER-PFAM MEMER BLIMPS-BLOCKS ProfileScan NOTIFS BLIMPS-PRINTS BLIMPS-FRIMPS-PRINTS BLAST-DMO BLAST-DMO BLAST-DMO BLAST-DMO BLAST-DMO BLAST-DMO BLAST-DMO

WO 01/42288

PC1/US00/33382

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
36	74/2011C01	307	S87-T288-S193	N5	Transmembrane domains: L23-L33, F99-R118, G204-R228 Rhodopsin-like G-protein coupled receptor (rhodopsin-like) signatures: G41-Y287 G-protein coupled receptor signatures: K90-P129, R234-R260, T279-Q295 G-protein coupled receptor signature: F102-T148 Olfactory receptor signatures: V271-L282, T288-G102 Melanocortin receptor family: S51-L63 Rhodopsin-like GPCR superfamily: F26-T50, M59-K80, F104-T126, L140-R161, R199-L222, M236-R260, K269-Q295 Olfactory receptor F0000921: L186-L245 Olfactory receptor PD149631: L246-S303 G-protein coupled receptor: DM0013 S29710 15-301: L17-L301 G-protein coupled receptor: DM0013 P23275 17-306: L17-L301 G-protein coupled receptor: DM0013 P23266 17-306: L17-L301 G-protein coupled receptor: DM0013 F4789 20-309: E23-E301	HMMER-FPAM HMMER BLIMPS-BLOCKS ProfileScan BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-FPROM PROCK PROCK ELAST-DOMO ELAST-DOMO ELAST-DOMO ELAST-DOMO

WO 01/42288

PC/T/US00/33382

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
37	7472012CD1	314	T19 S210 S291	N5 N38	7 transmembrane receptor (rhodopsin family) signatures: G11-E29 G-protein coupled receptor signatures: T482-K298 K90-P129 G-protein coupled receptor signature: Y102-H147 G-protein coupled receptor signature: F110-A125 Olfactory receptor signatures: M59-K80, F171-S191, F238-G253, I274-L285, S291-K305 Rhodopsin-like GPCR superfamily: P26-L50, K59-K80, F104-I128, I199-I222, T237-R261, R272-K298 G-protein coupled receptor: D600013 P23266 L7-305; I17-K303 G-protein coupled receptor: D600013 P23274 L8-306; E22-K303 G-protein coupled receptor: D600013 S291071 L8-306; P21-G299 G-protein coupled receptor: D600013 P30955 L8-305; P21-K303 Olfactory receptor PD149621: T246-T310 Olfactory receptor PD000921: L166-L245	BLIIMS-BLOCKS ProfilesScan MOTIFS BLIIMS-PRINTS BLIIMS-PRINTS BLAST-DMO BLAST-DMO BLAST-DMO BLAST-DMO BLAST-PRODOM BLAST-PRODOM

WO 01/42288

PC/T/US00/33382

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
71	7472019CS1	927		55586519_v113_9s_7.nt	1	927
72	7472019CS1	1032		55708153_v113_9s_6.nt	1	1032
73	7472019CS1	972		5213020_v113_9s_4.nt	1	972
74	7472019CS1	900		5234927_v113_9s_2.nt	1	900
75	7472019CS1	945		5213020_v113_9s_3.nt	1	945
76	7472019CS1	945		5213020_v113_9s_4.nt	1	945
77	7472019CS1	833		5213020_v113_9s_4.nt	1	833
78	7472019CS1	1080		55708153_v113_9s_3.nt	1	1080

WO 01/42288

PC/T/US00/33382

Table 5

Polynucleotide SEQ ID NO.	Invoice Project ID	Representative Library
61	2200534CB1	BRAYT04
62	327582LCB1	PROSEPT05
63	3744167CB1	LUNG0127

WO 01/42288

PCT/US00/33382

Table 6

LIBRARY	VECTOR	Library Description
BRAY/XT04	PSFOR11	Library was constructed using RNA isolated from separate populations of human astrocytes stimulated for 4 to 6 hours with a combination of cytokines including IL-1. The RNA was pooled for polyA RNA isolation and library construction.
LUNGNOT27	P1NCY	Library was constructed using RNA isolated from lung tissue removed from a 17-year-old Hispanic female.
PROSEPT06	P1NCY	Library was constructed using RNA isolated from diseased prostatic tissue from a 46-year-old Caucasian male during a radical prostatectomy and lymph node excision. Pathology indicated adenofibrotic hyperplasia. Pathology for the associated tumor tissue indicated grade 2 (of 4) adenocarcinoma, Gleason grade 3+3. The patient presented with elevated prostate specific antigen (PSA), proteinuria, decreased renal function, and urinary frequency. Patient history included hemiparesis, depressive disorder, sleep apnea, psoriasis, mitral valve prolapse, cerebrovascular disease, benign hypertension, and impotence. Family history included benign hypertension, cerebrovascular disease, and colon cancer.

WO 01/42288

PC/T/US00/33382

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.	
ABLIPARACEL FDF	A Fast Dual Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) <i>J. Mol. Biol.</i> 215:403-410; Altschul, S.F. et al. (1997) <i>Nucleic Acids Res.</i> 25:3389-3402.	ESY: Probability value=1.0E-8 or less Full Length requirement: Probability value=1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tblastn, fastx, fasty, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) <i>Proc. Natl. Acad. Sci. USA</i> 85:2444-2448; Pearson, W.R. (1990) <i>Methods Enzymol.</i> 183:63-98; and Smith, T.F. and M.S. Waterman (1981) <i>Adv. Appl. Math.</i> 2:482-489.	ESY: fasta E value=1.0E-6 Assembled ESY: fasta identity=95% or greater and Match length=200 bases or greater; fasta E value=1.0E-8 or less Full Length requirement: fastx score=1.00 or greater
BLIMPS	A Blocks IMPROved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases in search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) <i>Nucleic Acids Res.</i> 19:6566-6572; Henikoff, J.G. and S. Henikoff (1996) <i>Method. Enzymol.</i> 266:88-105; and Altschul, T.K. et al. (1997) <i>J. Chem. Inf. Comput. Sci.</i> 37:417-424.	Probability value=1.0E-3 or less
EMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences such as PFAM.	Krogh, A. et al. (1994) <i>J. Mol. Biol.</i> 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) <i>Nucleic Acids Res.</i> 26:3300-3323; Durbin, R. et al. (1998) <i>Our World View in a Mustell</i> , Cambridge Univ. Press, pp. 1-350.	PFAM hit: Probability value=1.0E-3 or less Signal peptide hit: Score=0 or greater

WO 01/42288

PC17/US00/33382

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.	Grishkov, M. et al. (1988) CABIOS 4:61-66; Grishkov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality scores:GCC-specified "HGH" value for that particular Prosite motif. Generally, score=1,4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phil's Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater. Match length= 56 or greater
Conseq	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claretie, J.M. and S. Aude (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Poisson, B. and P. Ayoub (1994) J. Mol. Biol. 237:182-192; Poisson, B. and P. Ayoub (1996) Protein Sci. 5:363-371.	
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 Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page N21-59; Genetics Computer Group, Madison, WI.	

WO 01/42288

PCT/US00/33382

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:1-39,
 - 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:1-39,
 - c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and
 - 10 d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-39.
- 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:40-78.
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 01/42288

PCT/US00/33382

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:40-78,
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:40-78,
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 01/42288

PCT/US00/33382

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-39.

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 01/42288

PCT/US00/33382

compound that specifically binds to the polypeptide of claim 1.

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

- 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.

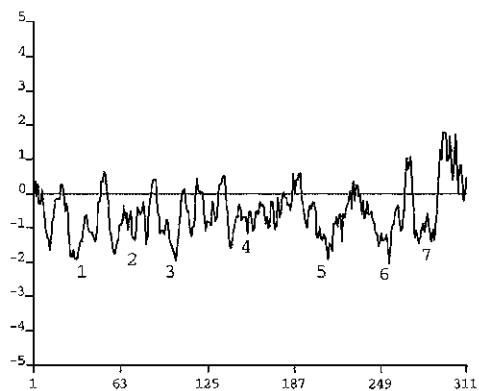
28. A method for assessing toxicity of a test compound, said method comprising:

- 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 01/42288

PCT/US00/33382

1/9



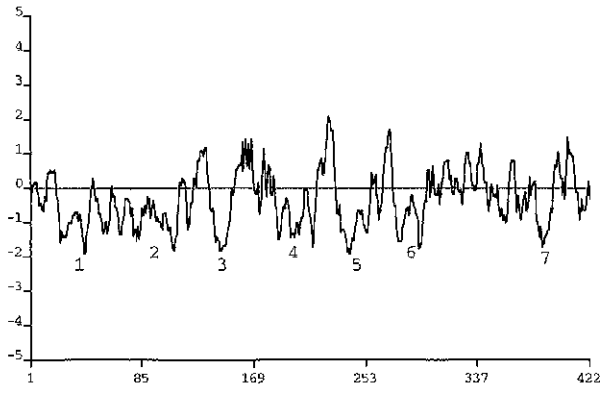
SEQ ID NO:1
(Incyte ID No. 104941CD1)

FIGURE 1

WO 01/42288

PCT/US00/33382

2/9



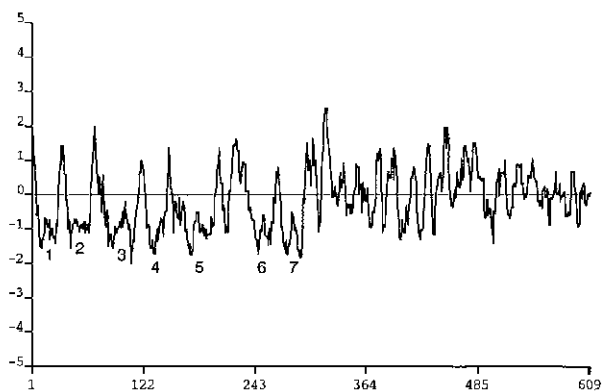
SEQ ID NO:3
(Incyte ID No. 3168839CD1)

FIGURE 2

WO 01/42288

PCT/US00/33382

3/9



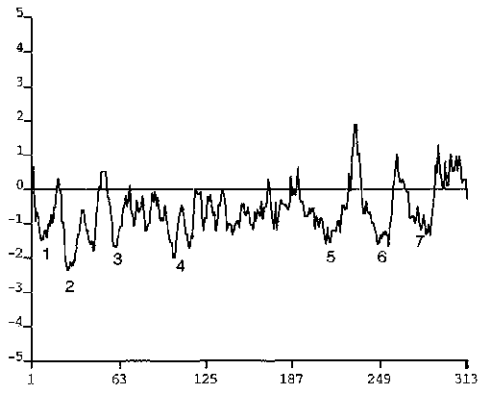
SEQ ID NO:4
(Incyte ID No. 3291235CD1)

FIGURE 3

WO 01/42288

PCT/US00/33382

4/9



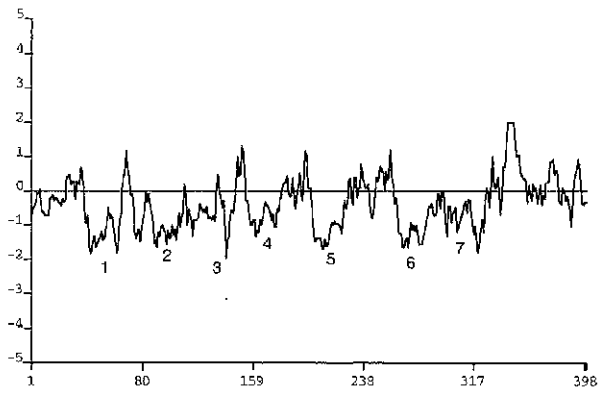
SEQ ID NO:5
(Incyte ID No. 7472001CD1)

FIGURE 4

WO 01/42288

PCT/US00/33382

5/9



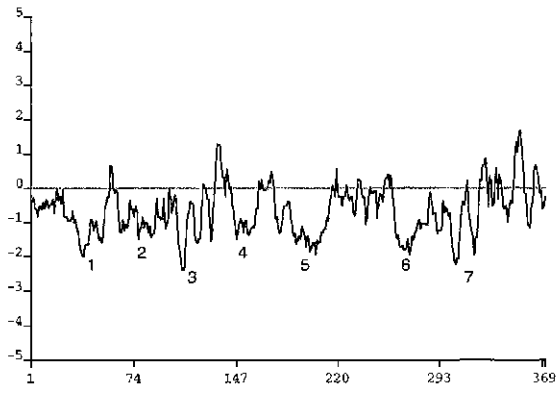
SEQ ID NO:6
(Incyte ID No. 7472003CD1)

FIGURE 5

WO 01/42288

PCT/US00/33382

6/9



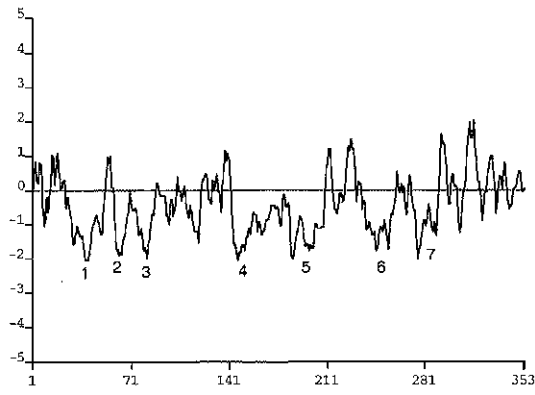
SEQ ID NO:7
(Incyte ID No. 7472004CD1)

FIGURE 6

WO 01/42288

PCT/US00/33382

7/9



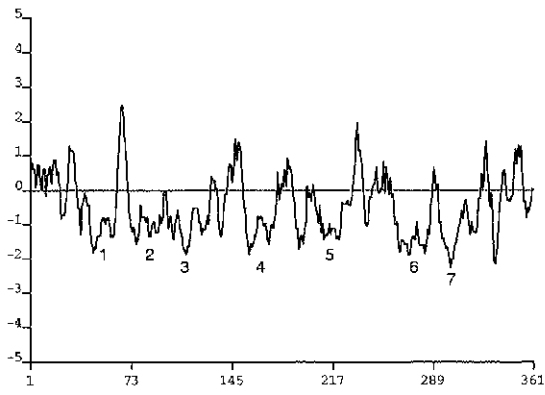
SEQ ID NO:19
(Incyte ID No. 3068234CD1)

FIGURE 7

WO 01/42288

PCT/US00/33382

8/9



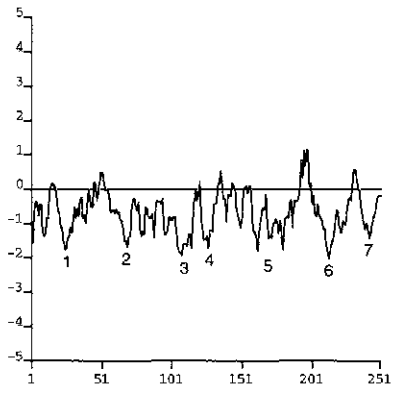
SEQ ID NO:20
(Incyte ID No. 5029478CD1)

FIGURE 8

WO 01/42288

PCT/US00/33382

9/9



SEQ ID NO:21
(Incyte ID No. 5182576CD1)

FIGURE 9

WO 01/42288

PCT/US00/33382

<110> INCYTE GENOMICS, INC.
BURFORD, Neil
BAUGHN, Mariah R.
AD-YOUNG, Janice
YANG, Junming
LI, Dzung Aina K.
REDDY, Roopa

<120> G-PROTEIN COUPLED RECEPTORS

<130> PI-0001 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/172,852; 60/171,732; 60/176,148; 60/177,331

<151> 1999-12-10; 1999 12-22; 2000-01-14; 2000-01-21

<160> 78

<170> PIRL Program

<210> 1

<211> 311

<212> PRE

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 104941CD1

<400> 1

```

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

```

WO 01/42288

PCT/US00/33382

Arg	Asp	Arg	Val	Ala	Thr	Val	Met	Tyr	Thr	Val	Val	Thr	Pro	Met
				275					280					285
Leu	Asn	Pro	Phe	Ile	Tyr	Ser	Leu	Arg	Asn	Lys	Asp	Met	Lys	Arg
				290					295					300
Gly	Leu	Lys	Lys	Leu	Gln	Asp	Arg	Ile	Tyr	Arg				
				305					310					

<210> 2
 <211> 891
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte TD No. 1499408CD1

<400> 2
 Met Asp Gln Pro Glu Ala Pro Cys Ser Ser Thr Gly Pro Arg Leu
 1 5 10 15
 Ala Val Ala Arg Glu Leu Leu Leu Ala Ala Leu Glu Glu Leu Ser
 20 25 30
 Gln Glu Gln Leu Lys Arg Phe Arg His Lys Leu Arg Asp Val Gly
 35 40 45
 Pro Asp Gly Arg Ser Ile Pro Trp Gly Arg Leu Glu Arg Ala Asp
 50 55 60
 Ala Val Asp Leu Ala Glu Gln Leu Ala Gln Phe Tyr Gly Pro Glu
 65 70 75
 Pro Ala Leu Glu Val Ala Arg Lys Thr Leu Lys Arg Ala Asp Ala
 80 85 90
 Arg Asp Val Ala Ala Gln Leu Gln Glu Arg Arg Leu Gln Arg Leu
 95 100 105
 Gly Leu Gly Ser Gly Thr Leu Leu Ser Val Ser Glu Tyr Lys Lys
 110 115 120
 Lys Tyr Arg Glu His Val Leu Gln Leu His Ala Arg Val Lys Glu
 125 130 135
 Arg Asn Ala Arg Ser Val Lys Ile Thr Lys Arg Phe Thr Lys Leu
 140 145 150
 Leu Ile Ala Pro Glu Ser Ala Ala Pro Glu Glu Ala Leu Gly Pro
 155 160 165
 Ala Glu Glu Pro Glu Pro Gly Arg Ala Arg Arg Ser Asp Thr His
 170 175 180
 Thr Phe Asn Arg Leu Phe Arg Arg Asp Glu Glu Gly Arg Arg Pro
 185 190 195
 Leu Thr Val Val Leu Gln Gly Pro Ala Gly Ile Gly Lys Thr Met
 200 205 210
 Ala Ala Lys Lys Ile Leu Tyr Asp Trp Ala Ala Gly Lys Leu Tyr
 215 220 225
 Gln Gly Gln Val Asp Phe Ala Phe Phe Met Pro Cys Gly Glu Leu
 230 235 240
 Leu Glu Arg Pro Gly Thr Arg Ser Leu Ala Asp Leu Ile Leu Asp
 245 250 255
 Gln Cys Pro Asp Arg Gly Ala Pro Val Pro Gln Met Leu Ala Gln
 260 265 270
 Pro Gln Arg Leu Leu Phe Ile Leu Asp Gly Ala Asp Glu Leu Pro
 275 280 285
 Ala Leu Gly Gly Pro Glu Ala Ala Pro Cys Thr Asp Pro Phe Glu
 290 295 300
 Ala Ala Ser Gly Ala Arg Val Leu Gly Gly Leu Leu Ser Lys Ala
 305 310 315
 Leu Leu Pro Thr Ala Leu Leu Leu Val Thr Thr Arg Ala Ala Ala
 320 325 330
 Pro Gly Arg Leu Gln Gly Arg Leu Cys Ser Pro Gln Cys Ala Glu
 335 340 345
 Val Arg Gly Phe Ser Asp Lys Asp Lys Lys Lys Tyr Phe Tyr Lys
 350 355 360
 Phe Phe Arg Asp Glu Arg Arg Ala Glu Arg Ala Tyr Arg Phe Val
 365 370 375

WO 01/42288

PCT/US00/33382

Lys Glu Asn Glu Thr Leu Phe Ala Leu Cys Phe Val Pro Phe Val 380 385 390
 Cys Trp Ile Val Cys Thr Val Leu Arg Gln Gln Leu Glu Leu Gly 395 400 405
 Arg Asp Leu Ser Arg Thr Ser Lys Thr Thr Thr Ser Val Tyr Leu 410 415 420
 Leu Phe Ile Thr Ser Val Leu Ser Ser Ala Pro Val Ala Asp Gly 425 430 435
 Pro Arg Leu Gln Gly Asp Leu Arg Asn Leu Cys Arg Leu Ala Arg 440 445 450
 Glu Gly Val Leu Gly Arg Arg Ala Gln Phe Ala Glu Lys Glu Leu 455 460 465
 Glu Gln Leu Glu Leu Arg Gly Ser Lys Val Gln Thr Leu Phe Leu 470 475 480
 Ser Lys Lys Glu Leu Pro Gly Val Leu Glu Thr Glu Val Thr Tyr 485 490 495
 Gln Phe Ile Asp Gln Ser Phe Gln Glu Phe Leu Ala Ala Leu Scr 500 505 510
 Tyr Leu Leu Glu Asp Gly Gly Val Pro Arg Thr Ala Ala Gly Gly 515 520 525
 Val Gly Thr Leu Leu Arg Gly Asp Ala Gln Pro His Ser His Leu 530 535 540
 Val Leu Thr Thr Arg Phe Leu Phe Gly Leu Leu Ser Ala Glu Arg 545 550 555
 Met Arg Asp Ile Glu Arg His Phe Gly Cys Met Val Ser Glu Arg 560 565 570
 Val Lys Gln Glu Ala Leu Arg Trp Val Gln Gly Gln Gly Gln Gly 575 580 585
 Cys Pro Gly Val Ala Pro Glu Val Thr Glu Gly Ala Lys Gly Leu 590 595 600
 Glu Asp Thr Glu Glu Pro Glu Glu Glu Glu Glu Cly Glu Glu Pro 605 610 615
 Asn Tyr Pro Leu Glu Leu Leu Tyr Cys Leu Tyr Glu Thr Gln Glu 620 625 630
 Asp Ala Phe Val Arg Gln Ala Leu Cys Arg Phe Pro Glu Leu Ala 635 640 645
 Leu Gln Arg Val Arg Phe Cys Arg Met Asp Val Ala Val Leu Scr 650 655 660
 Tyr Cys Val Arg Cys Cys Pro Ala Gly Gln Ala Leu Arg Leu Ile 665 670 675
 Ser Cys Arg Leu Val Ala Ala Gln Glu Lys Lys Lys Lys Ser Leu 680 685 690
 Gly Lys Arg Leu Gln Ala Ser Leu Gly Gly Gly Ser Ser Gln Gly 695 700 705
 Thr Thr Lys Gln Leu Pro Ala Ser Leu Leu His Pro Leu Phe Gln 710 715 720
 Ala Met Thr Asp Pro Leu Cys His Leu Ser Ser Leu Thr Leu Scr 725 730 735
 His Cys Lys Leu Pro Asp Ala Val Cys Arg Asp Leu Ser Glu Ala 740 745 750
 Leu Arg Ala Ala Pro Ala Leu Thr Glu Leu Gly Leu Leu His Asn 755 760 765
 Arg Leu Ser Glu Ala Gly Leu Arg Met Leu Ser Glu Gly Leu Ala 770 775 780
 Trp Pro Gln Cys Arg Val Gln Thr Val Arg Val Gln Leu Pro Asp 785 790 795
 Pro Gln Arg Gly Leu Gln Tyr Leu Val Gly Met Leu Arg Gln Ser 800 805 810
 Pro Ala Leu Thr Thr Leu Asp Leu Ser Cly Cys Gln Leu Pro Ala 815 820 825
 Pro Met Val Thr Tyr Leu Cys Ala Val Leu Gln His Gln Gly Cys 830 835 840
 Gly Leu Gln Thr Leu Ser Leu Ala Ser Val Glu Leu Ser Glu Gln 845 850 855
 Ser Leu Gln Glu Leu Gln Ala Val Lys Arg Ala Lys Pro Asp Leu 860 865 870
 Val Ile Thr His Pro Ala Leu Asp Gly His Pro Gln Pro Pro Lys

WO 01/42288 PCT/US00/33382

875 880 885

Glu Leu Ile Ser Thr Phe
896

<210> 3
<211> 422
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 3168833CD1

<400> 3

Met Leu Ala Asn Ser Ser Ser Thr Asn Ser Ser Val Leu Pro Cys
1 5 10 15
Pro Asp Tyr Arg Pro Thr His Arg Leu His Leu Val Val Tyr Ser
20 25 30
Leu Val Leu Ala Ala Gly Leu Pro Leu Asn Ala Leu Ala Leu Trp
35 40 45
Val Phe Leu Arg Ala Leu Arg Val His Ser Val Val Ser Val Tyr
50 55 60
Met Cys Asn Leu Ala Ala Ser Asp Leu Leu Phe Thr Leu Ser Leu
65 70 75
Pro Val Arg Leu Ser Tyr Tyr Ala Leu His His Trp Pro Phe Pro
80 85 90
Asp Leu Leu Cys Gln Thr Thr Gly Ala Ile Phe Gln Met Asn Met
95 100 105 110 115 120
Tyr Gly Ser Cys Ile Phe Leu Met Leu Ile Asn Val Asp Arg Tyr
125 130 135
Ala Ala Ile Val His Pro Leu Arg Leu Arg His Leu Arg Arg Pro
140 145 150 155 160 165
Arg Val Ala Arg Leu Leu Cys Leu Gly Val Trp Ala Leu Ile Leu
170 175 180 185 190 195
Val Phe Ala Val Pro Ala Ala Arg Val His Arg Pro Ser Arg Cys
200 205 210 215 220 225
Ser Gly Arg Val Phe Trp Thr Leu Ala Arg Pro Asp Ala Thr Gln
230 235 240 245 250 255
Ser Gln Arg Arg Arg Lys Thr Val Arg Leu Leu Leu Ala Asn Leu
260 265 270 275 280 285
Val Ile Phe Leu Leu Cys Phe Val Pro Tyr Asn Ser Thr Leu Ala
290 295 300 305 310 315
Val Tyr Gly Leu Leu Arg Ser Lys Leu Val Ala Ala Ser Val Trp
320 325 330 335 340 345
Ala Arg Asp Arg Val Arg Gly Val Leu Met Val Met Val Leu Leu
350 355 360 365 370 375
Ala Glu Gly Phe Arg Asn Thr Leu Arg Gly Leu Gly Thr Pro His
380 385 390 395 400 405
Arg Ala Arg Thr Ser Ala Thr Asn Gly Thr Arg Ala Ala Leu Ala
410 415 420 425 430 435
Gln Ser Glu Arg Ser Ala Val Thr Thr Asp Ala Thr Arg Pro Asp
440 445 450 455 460 465
Ala Ala Met Ser Pro Gly Phe Arg Pro Leu Asn Thr His Ala Ile
470 475 480 485 490 495
Ala Leu Ser Val Pro Asp Ser Gln Arg Leu Ser Phe Trp Glu Ala
500 505 510 515 520 525
Tyr Arg Val Tyr Thr Gln Glu Gly Gly Leu Gly Thr Trp Thr Phe
530 535 540 545 550 555
Gly Trp Gln Phe Gln Leu Ser Asn Ala Glu Glu Tyr Lys Val Trp

WO 01/42288

PCT/US00/33382

```

380                               385                               390
Pro Thr Arg Arg Phe Ser His Asp Asp Ala Asp Val Trp Ala Ala
395                               400                               405
Val Pro Leu Pro Ala Phe Leu Pro Arg Trp Gly Ser Gly Lys Asp
410                               415                               420
Leu Ser Ala Leu Ala His Leu Val Leu Pro Ala Gly Pro Glu Arg
425                               430                               435
Pro Arg Ala Ser Leu Leu Ala Phe Ala Glu Asp Ala Pro Leu Ser
440                               445                               450
Arg Ala Arg Arg Arg Ser Ala Glu Ser Leu Leu Ser Leu Arg Pro
455                               460                               465
Ser Ala Val Asp Ser Gly Pro Arg Gly Ala Arg Asp Ser Pro Pro
470                               475                               480
Gly Ser Pro Arg Arg Pro Gly Pro Gly Pro Arg Ser Ala Ser
485                               490                               495
Ala Ser Leu Leu Pro Asp Ala Phe Ala Leu Thr Ala Phe Glu Cys
500                               505                               510
Glu Pro Gln Ala Leu Arg Arg Pro Pro Gly Pro Phe Pro Ala Ala
515                               520                               525
Pro Ala Ala Pro Asp Gly Ala Asp Pro Gly Glu Ala Pro Thr Pro
530                               535                               540
Pro Ser Ser Ala Gln Arg Ser Pro Gly Pro Arg Pro Ser Ala His
545                               550                               555
Ser His Ala Gly Ser Leu Arg Pro Gly Leu Ser Ala Ser Trp Gly
560                               565                               570
Glu Pro Gly Gly Leu Arg Ala Ala Gly Gly Gly Ser Thr Ser
575                               580                               585
Ser Phe Leu Ser Ser Pro Ser Glu Ser Ser Gly Tyr Ala Thr Leu
590                               595                               600
His Ser Asp Ser Leu Gly Ser Ala Ser
605

```

```

<210> 5
<211> 313
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472001CD1

```

```

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

```

WO 01/42288

PC/T/US00/33382

```

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

```

```

<210> 6
<211> 398
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No. 7472003CD1

```

```

<400> 6
Met His Thr Val Ala Thr Ser Gly Pro Asn Ala Ser Trp Gly Ala
1      5      10      15
Pro Ala Asn Ala Ser Gly Cys Pro Gly Cys Gly Ala Asn Ala Ser
20      25      30
Asp Gly Pro Val Pro Ser Pro Arg Ala Val Asp Ala Trp Leu Val
35      40      45
Pro Leu Phe Phe Ala Ala Leu Met Leu Leu Gly Leu Val Gly Asn
50      55      60
Ser Leu Val Ile Tyr Val Ile Cys Arg His Lys Pro Met Arg Thr
65      70      75
Val Thr Asn Phe Tyr Ile Ala Asn Leu Ala Ala Thr Asp Val Thr
80      85      90
Phe Leu Leu Cys Cys Val Pro Phe Thr Ala Leu Leu Tyr Pro Leu
95      100      105
Pro Gly Trp Val Leu Gly Asp Phe Met Cys Lys Phe Val Asn Tyr
110      115      120
Ile Gln Gln Val Ser Val Gln Ala Thr Cys Ala Thr Leu Thr Ala
125      130      135
Met Ser Val Asp Arg Trp Tyr Val Thr Val Phe Pro Leu Arg Ala
140      145      150
Leu His Arg Arg Thr Pro Arg Leu Ala Leu Ala Val Ser Leu Ser
155      160      165
Ile Trp Thr Gly Ser Ala Ala Val Ser Ala Pro Val Leu Ala Leu
170      175      180
His Arg Leu Ser Pro Gly Pro Arg Ala Tyr Cys Ser Glu Ala Phe
185      190      195
Pro Ser Arg Ala Leu Glu Arg Ala Phe Ala Leu Tyr Asn Leu Leu
200      205      210
Ala Leu Tyr Leu Leu Pro Leu Leu Ala Thr Cys Ala Cys Tyr Ala
215      220      225
Ala Met Leu Arg His Leu Gly Arg Val Ala Val Arg Pro Ala Pro
230      235      240
Ala Asp Ser Ala Leu Gln Gly Gln Val Leu Ala Glu Arg Ala Gly
245      250      255
Ala Val Arg Ala Tyr Val Ser Arg Leu Val Ala Ala Val Val Leu
260      265      270
Leu Phe Ala Ala Cys Trp Gly Pro Ile Gln Leu Phe Leu Val Leu
275      280      285
Gln Ala Leu Gly Pro Ala Gly Ser Trp His Pro Arg Ser Tyr Ala

```

WO 01/42288

PCT/US00/33382

Ala	Tyr	Ala	Leu	Lys	Thr	Trp	Ala	His	Cys	Met	Ser	Tyr	Ser	Asn	300
				305					310					315	
Ser	Ala	Leu	Asn	Pro	Leu	Leu	Tyr	Ala	Phe	Leu	Gly	Ser	His	Phe	320
				325					330					335	
Arg	Gln	Ala	Phe	Arg	Arg	Val	Cys	Pro	Cys	Ala	Pro	Arg	Arg	Pro	340
				345					350					355	
Arg	Arg	Pro	Arg	Arg	Pro	Gly	Pro	Ser	Asp	Pro	Ala	Ala	Pro	His	360
				365					370					375	
Ala	Glu	Leu	Leu	Arg	Leu	Gly	Ser	His	Pro	Ala	Pro	Ala	Arg	Ala	380
				385					390					395	
Gln	Lys	Pro	Gly	Ser	Ser	Gly	Leu	Ala	Ala	Arg	Gly	Leu	Cys	Val	395
				395											

<210> ?
 <211> 369
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc feature
 <222> Incyte ID No. 7472004CD1

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

WO 01/42288 PCT/US00/33382

305 310 315
 Leu Asn Pro Leu Ile Tyr Tyr Trp Arg Ile Lys Lys Phe His Asp
 320 325 330
 Ala Cys Leu Asp Met Met Pro Lys Ser Phe Lys Phe Leu Pro Gln
 335 340 345
 Leu Pro Gly His Thr Lys Arg Arg Ile Arg Pro Ser Ala Val Tyr
 350 355 360
 Val Cys Gly Glu His Arg Thr Val Val
 365

<210> 8
 <211> 194
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc feature
 <223> Incyte ID No: 7475687CP1

<400> 8
 Met Ala Tyr Asp Arg Tyr Leu Ala Ile Cys Gln Pro Leu Arg Tyr
 1 5 10 15
 Pro Val Leu Met Asn Gly Arg Leu Cys Thr Val Leu Val Ala Gly
 20 25 30
 Ala Trp Val Ala Gly Ser Met His Gly Ser Ile Gln Ala Thr Leu
 35 40 45
 Thr Phe Arg Leu Pro Tyr Cys Gly Pro Asn Gln Val Asp Tyr Phe
 50 55 60
 Ile Cys Asp Ile Pro Ala Val Leu Arg Leu Ala Cys Ala Asp Thr
 65 70 75
 Thr Val Asn Glu Leu Val Thr Phe Val Asp Ile Gly Val Val Ala
 80 85 90
 Ala Ser Cys Phe Met Leu Ile Leu Leu Ser Tyr Ala Asn Ile Val
 95 100 105
 Asn Ala Ile Leu Lys Ile Arg Thr Thr Asp Gly Arg Arg Arg Ala
 110 115 120
 Phe Ser Thr Cys Gly Ser His Leu Ile Val Val Thr Val Tyr Tyr
 125 130 135
 Val Pro Cys Ile Phe Ile Tyr Leu Arg Ala Gly Ser Lys Gly Pro
 140 145 150
 Leu Asp Gly Ala Ala Ala Val Phe Tyr Thr Val Val Thr Pro Leu
 155 160 165
 Leu Asn Pro Leu Ile Tyr Thr Leu Arg Asn Gln Glu Val Lys Ser
 170 175 180
 Ala Leu Lys Arg Ile Thr Ala Gly Gln Ala Asp Val Asn Asn
 185 190

<210> 9
 <211> 173
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7483029CP1

<400> 9
 Met Tyr Leu Val Thr Val Leu Gly Asn Leu Leu Ile Ile Leu Ala
 1 5 10 15
 Thr Ile Ser Asp Ser His Leu His Thr Pro Met Tyr Phe Phe Leu
 20 25 30
 Ser Asn Leu Ser Phe Ala Asp Ile Cys Phe Val Ser Thr Thr Val
 35 40 45
 Pro Lys Met Leu Val Asn Ile Glu Thr Gln Ser Arg Val Ile Thr
 50 55 60
 Tyr Ala Asp Cys Ile Thr Gln Met Cys Phe Phe Ile Leu Phe Val
 65 70 75

WO 01/42288 PCT/US00/33382

Val Leu Asp Ser Leu Leu Leu Thr Val Met Ala Tyr Asp Arg Phe 80 85 90
 Val Ala Ile Cys His Pro Leu His Tyr Thr Val Ile Met Asp Ser 95 100 105
 Trp Leu Cys Gly Leu Leu Val Leu Val Ser Trp Ile Val Ser Ile 110 115 120
 Leu Tyr Ser Leu Leu Gln Ser Ile Met Ala Leu Gln Leu Ser Phe 125 130 135
 Cys Thr Glu Leu Lys Ile Pro His Phe Phe Cys Glu Leu Asn Gln 140 145 150
 Val Ile His Leu Ala Cys Ser Asp Thr Phe Ile Asn Asp Met Met 155 160 165
 Met Asn Phe Thr Ser Val Leu Leu 170

<210> 10
 <211> 220
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7477333CP1

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

<210> 11
 <211> 302
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7475164CP1

<400> 11
 Ala Glu Phe Ile Leu Ala Gly Leu Thr Gln Arg Pro Glu Leu Gln

WO 01/42288

PCT/US00/33382

```

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

```

```

<210> 12
<211> 110
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7473909CPI

```

```

<400> 12
Gly Pro Arg Thr Ala Ser Gly Cys Val Ile Met Ile Cys Phe Ala
1           5           10           15
Leu Thr Val Leu Ser Tyr Ile Arg Ile Leu Ala Thr Val Val Gln
20          25          30
Ile Arg Ser Ala Ala Ser Arg Arg Lys Ala Phe Ser Thr Cys Ser
35          40          45
Ser His Leu Gly Met Val Leu Leu Phe Tyr Gly Thr Gly Ser Ser
50          55          60
Thr Tyr Met Arg Pro Thr Thr Arg Tyr Ser Pro Leu Glu Gly Arg
65          70          75
Leu Ala Ala Val Phe Tyr Ser Ile Leu Ile Pro Thr Leu Asn Pro
80          85          90
Leu Ile Tyr Ser Leu Arg Asn Gln Asp Met Lys Arg Ala Leu Trp
95          100         105
Lys Leu Tyr Leu Gln

```

WO 01/42288

PCT/US00/33382

110

<210> 13
 <211> 178
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7475252CP1

<400> 13
 Glu Pro Glu Asn Leu Thr Gly Val Leu Glu Phe Leu Leu Leu Gly
 1 5 10 15
 Leu Pro Asp Asp Pro Glu Leu Gln Pro Val Leu Phe Gly Leu Phe
 20 25 30
 Leu Ser Met Tyr Leu Val Met Val Leu Gly Asn Leu Leu Ile Ile
 35 40 45
 Leu Ala Val Ser Ser Asp Ser His Leu His Ser Pro Met Tyr Phe
 50 55 60
 Phe Leu Ser Asn Leu Ser Leu Ala Asp Ile Gly Phe Ala Ser Thr
 65 70 75
 Thr Val Pro Lys Met Ile Val Asp Ile Gln Ala His Ser Arg Leu
 80 85 90
 Ile Ser Tyr Val Gly Cys Leu Thr Gln Met Ser Phe Leu Ile Phe
 95 100
 Phe Ala Cys Met Glu Ser Leu Leu Leu Ile Val Met Ala Tyr Asp
 105 110 115
 Arg Phe Val Ala Ile Cys His Pro Leu His Tyr Gln Val Ile Met
 120 125 130 135
 Ser Pro Arg Leu Cys Gly Phe Leu Val Leu Val Ser Phe Phe Leu
 140 145 150
 Ser Leu Leu Asp Ser Gln Leu His Asn Leu Ile Val Leu Glu Leu
 155 160 165
 Thr Cys Phe Asn Asp Val Glu Ile Ser Asn Phe Phe Leu
 170 175

<210> 14
 <211> 92
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7927572CP1

<400> 14
 Leu Leu Asp Ala Gln Leu Tyr Asn Leu Ile Ala Leu Gln Met Thr
 1 5 10 15
 Cys Phe Lys Asp Val Glu Ile Pro Asn Phe Phe Cys Asp Pro Ser
 20 25 30
 Gln Leu Pro His Leu Ala Cys Cys Asp Thr Phe Asn Asn Asn Ile
 35 40 45
 Ile Leu Tyr Phe Pro Asp Ala Ile Phe Gly Phe Leu Pro Ile Ser
 50 55 60
 Gly Thr Leu Phe Ser Tyr Asp Lys Ile Val Ser Ser Ile Leu Arg
 65 70 75
 Val Ser Ser Ser Gly Gly Lys Tyr Lys Ala Phe Ser Thr Tyr Gly
 80 85 90
 Ser His

<210> 15
 <211> 97
 <212> PRT
 <213> Homo sapiens

WO 01/42288

PC/T/US00/33382

```

<220>
<221> misc_feature
<223> Incyte ID No: 7481257CP1

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

```

```

<210> 16
<211> 133
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7485790CP1

<400> 16
Asp Pro Glu Leu Gln Pro Ile Leu Ala Gly Leu Ser Leu Ser Met
1 5 10 15
Tyr Leu Val Thr Val Leu Arg Asn Leu Leu Ile Ser Leu Ala Val
20 25 30
Ser Ser Asp Ser His Leu His Thr Pro Met Cys Phe Phe Leu Ser
35 40 45
Asn Leu Cys Trp Ala Asp Ile Gly Phe Thr Ser Ala Thr Val Pro
50 55 60
Lys Met Ile Val Asp Met Arg Ser His Ser Gly Val Ile Ser Tyr
65 70 75
Ala Asp Cys Leu Thr Arg Met Ser Phe Leu Val Leu Phe Ala Cys
80 85 90
Val Glu Asp Met Leu Leu Thr Val Met Ala Tyr Asp Cys Phe Val
95 100 105
Ala Ile Cys Arg Pro Leu His Tyr Pro Val Ile Val Asn Pro His
110 115 120
Leu Cys Val Phe Leu Val Ser Val Ser Phe Ser Leu Ala
125 130

```

```

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

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7482993CP1

<400> 17
Gly Ser Glu Cys Leu Leu Leu Ala Ala Met Ala Tyr Asp Arg Tyr
1 5 10 15
Ile Ala Ile Cys Asn Pro Leu Arg Tyr Ser Val Ile Leu Ser Lys
20 25 30
Val Leu Cys Asp Gln Leu Ala Ala Ser Cys Trp Ala Ala Gly Phe
35 40 45
Leu Asn Ser Val Val His Thr Val Leu Thr Phe Cys Leu Pro Phe
50 55 60

```

WO 01/42288

PCT/US00/33382

Cys Gly Asn Asn Glu Ile Asn Tyr Phe Phe Cys Asp Ile Pro Pro
65 70 75
Leu Leu Ile Leu Ser Cys Gly Asn Thr Ser Val Asn Glu Leu Ala
80 85 90
Leu Leu Ser Thr Gly Val Phe Ile Gly Trp Thr Pro Phe Leu Cys
95 100 105
Ile Val Leu Ser Tyr Ile Cys Ile Ile Ser Thr Ile Leu Arg Ile
110 115 120
Gln Ser Ser Glu Gly Arg Arg Lys Ala Phe Ser Thr Cys Ala Ser
125 130 135
His Leu Ala Ile Val Phe Leu Phe Tyr Gly Ser Ala Ile Phe Thr
140 145 150
Tyr Val Arg Pro Ile Ser Thr Tyr Ser Leu Lys Lys Asp Arg Leu
155 160 165
Val Ser Val Leu Tyr Ser Val Val Thr Pro Met Leu Asn Pro Ile
170 175 180
Ile Tyr Thr Leu Arg Asn Lys Asp Ile Lys Glu Ala Val Lys Thr
185 190 195
Ile Gly Ser Lys Trp Gln Pro Pro Ile Ser Ser Leu Asp Ser
200 205 210
Leu Thr Tyr

<210> 18
<211> 180
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No. 2829053CD1

<400> 18
Met Ser Glu Ala Ala Thr Arg Trp Ser Cys Gln Gly Ser Cys Gln
1 5 10 15
Lys Thr Cys Phe Ser Arg Val Arg Pro Trp Arg Arg Arg Cys Ser
20 25 30
Cys Gly Asp Ser Ser Ser Arg Arg Arg Arg Ser Cys Cys Thr Gly
35 40 45
Ser Leu Gly Pro Met Pro Arg Leu Pro Ser Leu Trp Pro Leu Ser
50 55 60
Leu Pro Leu Arg Ser Leu Ser Ser Pro His Arg Val Gln Gly Leu
65 70 75
Gly Pro Pro Arg Arg Leu Lys Ser Gln Leu Leu Pro Arg Phe Phe
80 85 90
Trp Arg Arg Gln Gln Glu Pro Leu Ser Ser Phe Pro Gly Arg Asn
95 100 105
Glu Gly Gly Ser Glu Met Glu Ile Leu Gly Val Cys Pro Val Ser
110 115 120
Pro Gly Ala Leu Ser Tyr Met Glu Ser Pro Thr Gly Phe Trp Arg
125 130 135
Pro Arg Glu Ala Ser Ser Leu Glu Leu Ala Lys Gly Ile Ser Lys
140 145 150
Arg Arg His Phe Leu Pro Ala Pro Ala Leu Cys Pro Asn Pro Arg
155 160 165
Ser Ser Glu Ala Phe Pro Gly Ala Val Cys Val Thr Leu Ala Ile
170 175 180

<210> 19
<211> 353
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No. 3068234CD1

WO 01/42288

PCT/US00/33382

```

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

```

```

<210> 20
<211> 361
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 5029478CD1

```

```

<400> 20
Met Ser Pro Glu Cys Ala Arg Ala Ala Gly Asp Ala Pro Leu Arg
1 5 10 15
Ser Leu Glu Gln Ala Asn Arg Thr Arg Phe Pro Phe Phe Ser Asp
20 25 30
Val Lys Gly Asp His Arg Leu Val Leu Ala Ala Val Glu Thr Thr
35 40 45
Val Leu Val Leu Ile Phe Ala Val Ser Leu Leu Gly Asn Val Cys

```

WO 01/42288

PCT/US00/33382

```

50          55          60
Ala Leu Val Leu Val Ala Arg Arg Arg Arg Arg Gly Ala Thr Ala
65
Cys Leu Val Leu Asn Leu Phe Cys Ala Arg Leu Leu Phe Ile Ser
80
Ala Ile Pro Leu Val Leu Ala Val Arg Trp Thr Glu Ala Trp Leu
95
Leu Gly Pro Val Ala Cys His Leu Leu Phe Tyr Val Met Thr Leu
110
Ser Gly Ser Val Thr Ile Leu Thr Leu Ala Ala Val Ser Leu Glu
125
Arg Met Val Cys Ile Val His Leu Gln Arg Gly Val Arg Gly Pro
140
Gly Arg Arg Ala Arg Ala Val Leu Leu Ala Leu Ile Trp Gly Tyr
155
Ser Ala Val Ala Ala Leu Pro Leu Cys Val Phe Phe Arg Val Val
170
Pro Gln Arg Leu Pro Gly Ala Asp Gln Glu Ile Ser Ile Cys Thr
185
Leu Ile Trp Pro Thr Ile Pro Gly Glu Ile Ser Trp Asp Val Ser
200
Phe Val Thr Leu Asn Phe Leu Val Pro Gly Leu Val Ile Val Ile
215
Ser Tyr Ser Lys Ile Leu Gln Ile Thr Lys Ala Ser Arg Lys Arg
230
Leu Thr Val Ser Leu Ala Tyr Ser Glu Ser His Gln Ile Arg Val
245
Ser Gln Gln Asp Phe Arg Leu Phe Arg Tyr Leu Phe Leu Leu Met
260
Val Ser Phe Phe Ile Met Trp Ser Pro Ile Ile Ile Thr Ile Leu
275
Leu Ile Leu Ile Gln Asn Phe Lys Gln Asp Leu Val Ile Trp Pro
290
Ser Leu Phe Phe Trp Val Val Ala Phe Thr Phe Ala Asn Ser Ala
305
Leu Asn Pro Ile Leu Tyr Asn Met Thr Leu Cys Arg Asn Glu Trp
320
Lys Lys Ile Phe Cys Cys Phe Trp Phe Pro Glu Lys Gly Ala Ile
335
Leu Thr Asp Thr Ser Val Lys Arg Asn Asp Leu Ser Ile Ile Ser
350          355          360
Gly

```

```

<210> 21
<211> 351
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: S102576CD1

```

```

<400> 21
Met Tyr Leu Val Thr Val Leu Arg Asn Leu Phe Ser Ile Leu Ala
1 5 15
Val Ser Ser Asp Cys Pro Leu His Thr Pro Met Tyr Phe Phe Leu
20 25 30
Ser Asn Leu Cys Trp Pro Asp Ile Gly Phe Thr Ser Ala Met Val
35 40 45
Pro Lys Met Ile Val Asp Thr Gln Ser His Ser Arg Val Ile Ser
50 55 60
His Ala Gly Cys Leu Thr Gln Met Ser Phe Leu Leu Leu Val Ala
65 70 75
Cys Ile Glu Gly Met Leu Leu Thr Val Met Ala Tyr Asp Cys Phe
80 85 90
Val Ala Ile Cys Arg Pro Leu His Tyr Pro Val Ile Val Asn Pro

```

WO 01/42288

PCT/US00/33382

```

95 100 105
His Leu Cys Val Phe Phe Val Leu Val Ser Phe Phe Leu Ser Leu
110 115 120
Leu Asp Ser Gln Leu His Ser Trp Ile Val Leu Gln Leu Thr Ile
125 130 135
Ile Lys Asn Val Glu Ile Ser Asn Leu Val Cys Asp Pro Ser Gln
140 145 150
Leu Leu Asn Leu Ala Cys Ser Asp Ser Val Ile Asn Asn Ile Phe
155 160 165
Ile Tyr Phe Asp Ser Thr Met Phe Gly Phe Leu Pro Ile Ser Gly
170 175 180
Ile Phe Leu Ser Tyr Tyr Lys Ile Val Pro Ser Ile Leu Arg Ile
185 190 195
Ser Ser Ser Asp Gly Lys Tyr Lys Ala Phe Ser Thr Cys Gly Cys
200 205 210
His Leu Ala Val Val Cys Trp Phe Tyr Gly Thr Gly Ile Gly Met
215 220 225
Tyr Leu Thr Ser Ala Val Ser Pro Pro Pro Arg Asn Gly Val Val
230 235 240
Ala Ser Val Met Tyr Ala Val Val Thr Pro Cys
245 250

```

```

<210> 22
<211> 315
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 2209534CD1

```

```

<400> 22
Met Lys Ala Asn Tyr Ser Ala Glu Glu Arg Phe Leu Leu Leu Gly
1 15
Phe Ser Asp Trp Pro Ser Leu Gln Pro Val Leu Phe Ala Leu Val
20 25 30
Leu Leu Cys Tyr Leu Leu Thr Leu Thr Gly Asn Ser Ala Leu Val
35 40 45
Leu Leu Ala Val Arg Asp Pro Arg Leu His Thr Pro Met Tyr Tyr
50 55 60
Phe Leu Cys His Leu Ala Leu Val Asp Ala Gly Phe Thr Thr Ser
65 70 75
Val Val Pro Pro Leu Leu Ala Asn Leu Arg Gly Pro Ala Leu Trp
80 85 90
Leu Pro Arg Ser His Cys Thr Ala Gln Leu Cys Ala Ser Leu Ala
95 100 105
Leu Gly Ser Ala Gln Cys Val Leu Leu Ala Val Met Ala Leu Asp
110 115 120
Arg Ala Ala Ala Val Cys Arg Pro Leu Arg Tyr Ala Gly Leu Val
125 130 135
Ser Pro Arg Leu Cys Arg Thr Leu Ala Ser Ala Ser Trp Leu Ser
140 145 150
Gly Leu Thr Asn Ser Val Ala Gln Thr Ala Leu Leu Ala Glu Arg
155 160 165
Pro Leu Cys Ala Pro Arg Leu Leu Asp His Phe Ile Cys Gln Leu
170 175 180
Pro Ala Leu Leu Lys Leu Ala Cys Gly Gly Asp Gly Asp Thr Thr
185 190 195
Glu Asn Gln Met Phe Ala Ala Arg Val Val Ile Leu Leu Leu Pro
200 205 210
Phe Ala Val Ile Leu Ala Ser Tyr Gly Ala Val Ala Arg Ala Val
215 220 225
Cys Cys Met Arg Phe Ser Gly Gly Arg Arg Arg Ala Val Gly Thr
230 235 240
Cys Gly Ser His Leu Thr Ala Val Cys Leu Phe Tyr Gly Ser Ala
245 250 255
Ile Tyr Thr Tyr Leu Gln Pro Ala Gln Arg Tyr Asn Gln Ala Arg

```

WO 01/42288

PCT/US00/33382

Gly	Lys	Phe	Val	Ser	Leu	Phe	Tyr	Thr	Val	Val	Thr	Pro	Ala	Leu	260	265	270
Asn	Pro	Leu	Ile	Tyr	Thr	Leu	Arg	Asn	Lys	Lys	Val	Iys	Gly	Ala	275	280	285
Ala	Arg	Arg	Leu	Leu	Arg	Ser	Leu	Gly	Arg	Gly	Gln	Ala	Gly	Ala	290	295	300
															305	310	315

<210> 23
 <211> 470
 <212> PRI
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No. 3275821CD1

<400> 23
 Met Asp Thr Thr Met Glu Ala Asp Leu Gly Ala Thr Gly His Arg
 1 5 10 15
 Pro Arg Thr Glu Leu Asp Asp Glu Asp Ser Tyr Pro Gln Gly Gly
 20 25 30
 Trp Asp Thr Val Phe Leu Val Ala Leu Leu Leu Gly Leu Pro
 35 40 45
 Ala Asn Gly Leu Met Ala Trp Leu Ala Gly Ser Gln Ala Arg His
 50 55 60
 Gly Ala Gly Thr Arg Leu Ala Leu Leu Leu Ser Leu Ala Leu
 65 70 75
 Ser Asp Phe Leu Phe Leu Ala Ala Ala Ala Phe Gln Ile Leu Glu
 80 85 90
 Ile Arg His Gly Gly His Trp Pro Leu Gly Thr Ala Ala Cys Arg
 95 100 105
 Phe Tyr Tyr Phe Leu Trp Gly Val Ser Tyr Ser Ser Gly Leu Phe
 110 115 120
 Leu Leu Ala Ala Leu Ser Leu Asp Arg Cys Leu Leu Ala Leu Cys
 125 130 135
 Pro His Trp Tyr Pro Gly His Arg Pro Val Arg Leu Pro Leu Trp
 140 145 150
 Val Cys Ala Gly Val Trp Val Leu Ala Thr Leu Phe Ser Val Pro
 155 160 165
 Trp Leu Val Phe Pro Glu Ala Ala Val Trp Trp Tyr Asp Leu Val
 170 175 180
 Ile Cys Leu Asp Phe Trp Asp Ser Glu Glu Leu Ser Leu Arg Met
 185 190 195
 Leu Glu Val Leu Gly Gly Phe Leu Pro Phe Leu Leu Leu Val
 200 205 210
 Cys His Val Leu Thr Gln Ala Thr Ala Cys Arg Thr Cys His Arg
 215 220 225
 Gln Gln Gln Pro Ala Ala Cys Arg Gly Phe Ala Arg Val Ala Arg
 230 235 240
 Thr Ile Leu Ser Ala Tyr Val Val Leu Arg Leu Pro Tyr Gln Leu
 245 250 255
 Ala Gln Leu Leu Tyr Leu Ala Phe Leu Trp Asp Val Tyr Ser Gly
 260 265 270
 Tyr Leu Leu Trp Glu Ala Leu Val Tyr Ser Asp Tyr Leu Ile Leu
 275 280 285
 Leu Asn Ser Cys Leu Ser Pro Phe Leu Cys Leu Met Ala Ser Ala
 290 295 300
 Asp Leu Arg Thr Leu Leu Arg Ser Val Leu Ser Ser Phe Ala Ala
 305 310 315
 Ala Leu Cys Glu Glu Arg Pro Gly Ser Phe Thr Pro Thr Glu Pro
 320 325 330
 Gln Thr Gln Leu Asp Ser Glu Gly Pro Thr Leu Pro Glu Pro Met
 335 340 345
 Ala Glu Ala Gln Ser Gln Met Asp Pro Val Ala Gln Pro Gln Val
 350 355 360

WO 01/42288 PCT/US00/33382

Asn Pro Thr Leu Gln Pro Arg Ser Asp Pro Thr Ala Gln Pro Gln 365 370 375
 Leu Asn Pro Thr Ala Gln Pro Gln Ser Asp Pro Thr Ala Glu Pro 380 385 390
 Gln Leu Asn Leu Met Ala Gln Pro Gln Ser Asp Ser Val Ala Gln 395 400 405
 Pro Gln Ala Asp Thr Asn Val Gln Thr Pro Ala Pro Ala Ala Ser 410 415 420
 Ser Val Pro Ser Pro Cys Asp Glu Ala Ser Pro Thr Pro Ser Ser 425 430 435
 His Pro Thr Pro Gly Ala Leu Glu Asp Pro Ala Thr Pro Pro Ala 440 445 450
 Ser Glu Gly Glu Ser Pro Ser Ser Thr Pro Pro Glu Ala Ala Pro 455 460 465
 Gly Ala Gly Pro Thr 470

<210> 34
 <211> 358
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3744167CD1

<400> 24

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

WO 01/42288 PCT/US00/33382

Pro Arg Phe Leu Thr Arg Leu Phe Glu Gly Ser Gly Glu Ala Arg
 305 310 315
 Gly Gly Gly Arg Ser Arg Glu Gly Thr Met Glu Leu Arg Thr Thr
 320 325 330
 Pro Gln Leu Lys Val Val Gly Gln Gly Arg Gly Asn Gly Asp Pro
 335 340 345
 Gly Gly Gly Met Glu Lys Asp Gly Pro Glu Trp Asp Leu
 350 355

<210> 25
 <211> 314
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> Incyte ID No: 7472007CD1

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

<210> 26
 <211> 305
 <212> PRT
 <213> Homo sapiens

WO 01/42288

PC/T/US00/33382

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472008cd1

<400> 26
Met Glu Gly Ser Val Glu Ala Thr Pro Glu Ile Pro Ala Gln Met
1      5      10      15      20      25      30      35      40      45
Lys Cys His Pro Ser Arg Pro Ser Thr Leu Asn Gln Leu Ser Phe
20     25     30     35     40     45     50     55     60     65
Tyr Gly Ala Val Ser Ser Leu Gly Arg Met His Gly Leu Glu Thr
35     40     45     50     55     60     65     70     75     80
Lys Ser Ser Ala Glu Ile Arg Ala Gly Leu Lys Arg Cys Asp Thr
40     45     50     55     60     65     70     75     80     85
Leu Val Leu Glu Ala Ser Thr Leu Glu Cys Asn Met Val Ile Val
65     70     75     80     85     90     95     100    105    110
Leu Val Ser Leu Lys Asp Pro Lys Leu His Ile Pro Met Tyr Phe
80     85     90     95     100    105    110    115    120    125
Phe Leu Ser Asn Leu Ser Leu Val Asp Leu Cys Leu Thr Ser Ser
95     100   105   110   115   120   125   130   135   140
Cys Val Pro Gln Met Leu Ile Asn Phe Trp Gly Pro Glu Lys Thr
110    115   120   125   130   135   140   145   150   155
Ile Ser Tyr Ile Gly Cys Ala Ile Gln Leu Tyr Val Phe Leu Trp
125    130   135   140   145   150   155   160   165   170
Leu Gly Ala Thr Glu Tyr Val Leu Leu Val Val Met Ala Val Asp
140    145   150   155   160   165   170   175   180   185
Cys Tyr Val Ala Val Cys His Pro Leu Gln Asn Thr Met Ile Met
145    150   155   160   165   170   175   180   185   190
His Pro Lys Leu Cys Leu Gln Leu Ala Ile Leu Ala Trp Gly Thr
170    175   180   185   190   195   200   205   210   215
Gly Leu Ala Gln Ser Leu Ile Gln Ser Pro Ala Thr Leu Arg Leu
185    190   195   200   205   210   215   220   225   230
Pro Phe Cys Ser Gln Arg Met Val Asp Asp Val Val Cys Glu Val
200    205   210   215   220   225   230   235   240   245
Pro Ala Leu Ile Gln Leu Ser Ser Thr Asp Thr Thr Tyr Ser Glu
215    220   225   230   235   240   245   250   255   260
Ile Gln Met Ser Ile Ala Ser Val Val Leu Leu Val Met Pro Leu
230    235   240   245   250   255   260   265   270   275
Ile Ile Ile Leu Ser Ser Ser Gly Ala Ile Ala Lys Ala Val Leu
245    250   255   260   265   270   275   280   285   290
Arg Ile Lys Ser Thr Ala Gly Gln Lys Lys Ala Phe Gly Thr Cys
260    265   270   275   280   285   290   295   300   305
Ile Ser His Leu Leu Val Val Ser Leu Phe Tyr Gly Thr Val Thr
275    280   285   290   295   300   305   310   315   320
Gly Val Tyr Leu Gln Pro Lys Asn His Tyr Pro His Glu Trp Gly
290    295   300   305   310   315   320   325   330   335
Lys Phe Leu Thr Leu Phe Tyr Thr Val Val Thr Pro Thr Leu Asn
305    310   315   320   325   330   335   340   345   350
Pro Leu Ile Tyr Thr Leu Arg Asn Lys Glu Leu His Pro Trp Leu
320    325   330   335   340   345   350   355   360
Lys Glu Ala Lys Val Gln Thr Ala Ser Glu Ser Ala Ser Pro Lys
335    340   345   350   355   360
His Trp Glu Leu Pro His Gly Val Gly Pro Val Gly Val Gln Lys
350    355   360
Thr Arg Thr Glu Leu
365

```

```

<210> 27
<211> 317
<212> FRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472013cd1

```

```

<400> 27
Met Ser Phe Ala Pro Asn Ala Ser His Ser Pro Val Phe Leu Leu

```

WO 01/42288

PCT/US00/33382

```

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

```

```

<210> 28
<211> 335
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472015CD1

```

```

<400> 28
Met Glu Ser Ser Phe Ser Phe Gly Val Ile Leu Ala Val Leu Ala
1           5           10           15
Ser Leu Ile Ile Ala Thr Asn Thr Leu Val Ala Val Ala Val Leu
20
Leu Leu Ile His Lys Asn Asp Gly Val Ser Leu Cys Phe Thr Leu
35
Asn Leu Ala Val Ala Asp Thr Leu Ile Gly Val Ala Ile Ser Gly
50
Leu Leu Thr Asp Gln Leu Ser Ser Pro Ser Arg Pro Thr Gln Lys
65
Thr Leu Cys Ser Leu Arg Met Ala Phe Val Thr Ser Ser Ala Ala
80
Ala Ser Val Leu Thr Val Met Leu Ile Thr Phe Asp Arg Tyr Leu

```

WO 01/42288

PCT/US00/33382

```

95      100      105
Ala Ile Lys Gln Pro Phe Arg Tyr Leu Lys Ile Met Ser Gly Phe
110      115      120
Val Ala Gly Ala Cys Ile Ala Gly Leu Trp Leu Val Ser Tyr Leu
125      130      135
Ile Gly Phe Leu Pro Leu Gly Ile Pro Met Phe Gln Gln Thr Ala
140      145      150
Tyr Lys Gly Gln Cys Ser Phe Phe Ala Val Phe His Pro His Phe
155      160      165
Val Leu Thr Leu Ser Cys Val Gly Phe Phe Pro Ala Met Leu Leu
170      175      180
Phe Val Phe Phe Tyr Cys Asp Met Leu Lys Ile Ala Ser Met His
185      190      195
Ser Gln Gln Ile Arg Lys Met Glu His Ala Gly Ala Met Ala Gly
200      205      210
Gly Tyr Arg Ser Pro Arg Thr Pro Ser Asp Phe Lys Ala Leu Arg
215      220      225
Thr Val Ser Val Leu Ile Gly Ser Phe Ala Leu Ser Trp Thr Pro
230      235      240
Phe Leu Ile Thr Gly Ile Val Gln Val Ala Cys Gln Glu Cys His
245      250      255
Leu Tyr Leu Val Leu Glu Arg Tyr Leu Trp Leu Leu Gly Val Gly
260      265      270
Asn Ser Leu Leu Asn Pro Leu Ile Tyr Ala Tyr Trp Gln Lys Glu
275      280      285
Val Arg Leu Gln Leu Tyr His Met Ala Leu Gly Val Lys Lys Val
290      295      300
Leu Thr Ser Phe Leu Leu Phe Leu Ser Ala Arg Asn Cys Gly Pro
305      310      315
Glu Arg Pro Arg Glu Ser Ser Cys His Ile Val Thr Ile Ser Ser
320      325      330
Ser Glu Phe Asp Gly
335

```

```

<210> 39
<211> 309
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> TrcYTE ID No: 7472016CD1

```

```

<400> 29
Met Arg Glu Asn Asn Gln Ser Ser Thr Leu Glu Phe Ile Leu Leu
1      5      10      15
Gly Val Thr Gly Gln Gln Glu Gln Glu Asp Phe Phe Tyr Ile Leu
20      25      30
Phe Leu Phe Ile Tyr Pro Ile Thr Leu Ile Gly Asn Leu Leu Ile
35      40      45
Val Leu Ala Ile Cys Ser Asp Val Arg Leu His Asn Pro Met Tyr
50      55      60
Phe Leu Leu Ala Asn Leu Ser Leu Val Asp Ile Phe Phe Ser Ser
65      70      75
Val Thr Ile Pro Lys Met Leu Ala Asn His Leu Leu Gly Ser Lys
80      85      90
Ser Ile Ser Phe Gly Gly Cys Leu Thr Gln Met Tyr Phe Met Ile
95      100      105
Ala Leu Gly Asn Thr Asp Ser Tyr Ile Leu Ala Ala Met Ala Tyr
110      115      120
Asp Arg Ala Val Ala Ile Ser His Pro Leu His Tyr Thr Thr Ile
125      130      135
Met Ser Pro Arg Ser Cys Ile Trp Leu Ile Ala Gly Ser Trp Val
140      145      150
Ile Gly Asn Ala Asn Ala Leu Pro His Thr Leu Leu Thr Ala Ser
155      160      165
Leu Ser Phe Cys Gly Asn Gln Glu Val Ala Asn Phe Tyr Cys Asp

```

WO 01/42288

PCT/US00/33382

```

170                               175                               180
Ile Thr Pro Leu Leu Lys Leu Ser Cys Ser Asp Ile His Phe His
185                               190                               195
Val Lys Met Met Tyr Leu Gly Val Gly Ile Phe Ser Val Pro Leu
200                               205                               210
Leu Cys Ile Ile Val Ser Tyr Ile Arg Val Phe Ser Thr Val Phe
215                               220                               225
Gln Val Pro Ser Thr Lys Gly Val Leu Lys Ala Phe Ser Thr Cys
230                               235                               240
Gly Ser His Leu Thr Val Val Ser Leu Tyr Tyr Gly Thr Val Met
245                               250                               255
Gly Thr Tyr Phe Arg Pro Leu Thr Asn Tyr Ser Leu Lys Asp Ala
260                               265                               270
Val Ile Thr Val Met Tyr Thr Ala Val Thr Pro Met Leu Asn Pro
275                               280                               285
Phe Ile Tyr Ser Leu Arg Asn Arg Asp Met Lys Ala Ala Leu Arg
290                               295                               300
Lys Leu Phe Asn Lys Arg Ile Ser Ser
305

```

```

<210> 30
<211> 256
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<222> Incyte ID No: 7472017CD1

```

```

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

```

```

<210> 31
<211> 263
<212> PRT
<213> Homo sapiens

```

WO 01/42288

PC/T/US00/33382

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472018CD1

<400> 31
Met Gly Pro Gly Glu Ala Leu Leu Ala Gly Leu Leu Val Met Val
1 5 10 15
Leu Ala Val Ala Leu Leu Ser Asn Ala Leu Val Leu Leu Cys Cys
20 25 30
Ala Tyr Ser Ala Glu Leu Arg Thr Arg Ala Ser Gly Val Leu Leu
35 40 45
Val Asn Leu Ser Leu Gly His Leu Leu Leu Ala Ala Leu Asp Met
50 55 60
Pro Phe Thr Leu Leu Gly Val Met Arg Gly Arg Thr Pro Ser Ala
65 70 75
Pro Gly Ala Cys Gln Val Ile Gly Phe Leu Asp Thr Phe Leu Ala
80 85 90
Ser Asn Ala Ala Leu Ser Val Ala Ala Leu Ser Ala Asp Gln Trp
95 100 105
Leu Ala Val Gly Phe Pro Leu Arg Tyr Ala Gly Arg Leu Arg Pro
110 115 120
Arg Tyr Ala Gly Leu Leu Leu Gly Cys Ala Trp Gly Gln Ser Leu
125 130 135
Ala Phe Ser Gly Ala Ala Leu Gly Cys Ser Trp Leu Gly Tyr Ser
140 145 150
Ser Ala Phe Ala Ser Cys Ser Leu Arg Leu Pro Pro Glu Pro Glu
155 160 165
Arg Pro Arg Phe Ala Ala Phe Thr Ala Thr Leu His Ala Val Gly
170 175 180
Phe Val Leu Pro Leu Ala Val Leu Cys Leu Thr Ser Leu Gln Val
185 190 195
His Arg Val Ala Arg Arg His Cys Gln Arg Met Asp Thr Val Thr
200 205 210
Met Lys Ala Leu Ala Leu Leu Ala Asp Leu His Pro Ser Val Arg
215 220 225
Gln Arg Cys Leu Ile Gln Gln Lys Arg Arg Arg His Arg Ala Thr
230 235 240
Arg Lys Ile Gly Ile Ala Ile Ala Thr Phe Leu Ile Cys Phe Ala
245 250 255
Pro Tyr Val Met Thr Arg Leu Ala Glu Leu Val Pro Phe Val Thr
260 265 270
Val Asn Ala Gln Trp Gly Ile Leu Ser Lys Cys Leu Thr Tyr Ser
275 280 285
Lys Ala Val Ala Asp Pro Phe Thr Tyr Ser Leu Leu Arg Arg Pro
290 295 300
Phe Arg Gln Val Leu Ala Gly Met Val His Arg Leu Leu Lys Arg
305 310 315
Thr Pro Arg Pro Ala Ser Thr His Asp Ser Ser Leu Asp Val Ala
320 325 330
Gly Met Val His Gln Leu Leu Lys Arg Thr Pro Arg Pro Ala Ser
335 340 345
Thr His Asn Gly Ser Val Asp Thr Glu Asn Asp Ser Cys Leu Gln
350 355 360
Gln Thr His

```

```

<210> 32
<211> 308
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472019CD1

```

```

<400> 32
Met Ala Met Asp Asn Val Thr Ala Val Phe Gln Phe Leu Leu Ile

```

WO 01/42288

PCT/US00/33382

```

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

```

```

<210> 33
<211> 343
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<222> Incyte ID No: 7472021CD1

```

```

<400> 33
Met His Phe Leu Pro Thr Val Phe Gly Phe Leu Asn Arg Val Thr
1           5           10           15
Leu Gly Ile Phe Arg Glu Thr Met Val Asn Leu Thr Ser Met Ser
20          25          30
Gly Phe Leu Leu Met Gly Phe Ser Asp Glu Arg Lys Leu Gln Ile
35          40          45
Leu His Ala Leu Val Phe Leu Val Thr Tyr Leu Leu Ala Leu Thr
50          55          60
Gly Asn Leu Leu Ile Ile Thr Ile Ile Thr Val Asp Arg Arg Leu
65          70          75
His Ser Pro Met Tyr Tyr Phe Leu Lys His Leu Ser Leu Leu Asp
80          85          90
Leu Cys Phe Ile Ser Val Thr Val Pro Gln Ser Ile Ala Asn Ser
95          100         105
Leu Met Gly Asn Gly Tyr Ile Ser Leu Val Gln Cys Ile Leu Gln

```

WO 01/42288

PC/T/US00/33382

```

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

```

```

<210> 34
<211> 323
<212> FRP
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472009CD1

```

```

<400> 34
Met Trp Gln Lys Asn Gln Thr Ser Leu Ala Asp Phe Ile Leu Glu
1          5          10          15
Gly Leu Phe Asp Asp Ser Leu Thr His Leu Phe Leu Phe Ser Leu
20          25          30
Thr Met Val Val Phe Leu Ile Ala Val Ser Gly Asn Thr Leu Thr
35          40          45
Ile Leu Leu Ile Cys Ile Asp Pro Gln Leu His Thr Pro Met Tyr
50          55          60
Phe Leu Leu Ser Gln Leu Ser Leu Met Arg Leu Met His Val Ser
65          70          75
Thr Thr Ile Leu Lys Met Ala Thr Asn Tyr Leu Ser Gly Lys Lys
80          85          90
Ser Ile Ser Phe Val Gly Cys Ala Thr Gln His Phe Leu Tyr Leu
95          100          105
Cys Leu Gly Gly Ala Glu Cys Phe Leu Leu Ala Val Met Ser Tyr
110          115          120
Asp Arg Tyr Val Ala Ile Cys His Pro Leu Arg Tyr Ala Val Leu
125          130          135
Met Asn Lys Lys Val Gly Leu Met Met Ala Val Met Ser Trp Leu
140          145          150
Gly Ala Ser Val Asn Ser Leu Ile His Met Ala Ile Leu Met His
155          160          165
Phe Pro Phe Cys Gly Pro Arg Lys Val Tyr His Phe Tyr Cys Gln
170          175          180
Phe Pro Ala Val Val Lys Leu Val Cys Gly Asp Ile Thr Val Tyr

```

WO 01/42288

PCT/US00/33382

```

185          190          195
Glu Thr Thr Val Tyr Ile Ser Ser Ile Leu Leu Leu Leu Pro Ile
200
Phe Leu Ile Ser Thr Ser Tyr Val Phe Ile Leu Gln Ser Val Ile
210
Gln Met Arg Ser Thr Gly Ser Lys Arg Asn Ala Phe Ala Thr Cys
220
Gly Ser His Leu Thr Val Val Ser Leu Trp Phe Gly Ala Cys Ile
230
Phe Ser Tyr Met Arg Pro Arg Ser Gln Cys Thr Leu Leu Gln Asn
240
Lys Val Gly Ser Val Phe Tyr Ser Ile Ile Thr Pro Thr Leu Asn
250
Ser Leu Ile Tyr Thr Leu Arg Asn Lys Asp Val Ala Lys Ala Leu
260
Arg Arg Val Leu Arg Arg Asp Val Ile Thr Gln Cys Ile Gln Arg
270
Leu Gln Leu Trp Leu Pro Arg Val
280

```

```

<210> 35
<211> 299
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472010CD1

```

```

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

```

WO 01/42288 PCT/US00/33382

275 280 285
 Asp Lys Leu Ala Leu Gly Lys Thr Leu Lys Arg Val Leu Phe
 290 295

<210> 36
 <211> 307
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <225> Incyte ID No: 7472011CD1

<400> 36
 Met Glu Thr Gly Asn Leu Thr Trp Val Ser Asp Phe Val Phe Leu
 1 5 10 15
 Gly Leu Ser Gln Thr Arg Glu Leu Gln Arg Phe Leu Phe Leu Met
 20 25 30
 Phe Leu Phe Val Tyr Ile Thr Thr Val Met Gly Asn Ile Leu Ile
 35 40 45
 Ile Ile Thr Val Thr Ser Asp Ser Gln Leu His Thr Pro Met Tyr
 50 55 60
 Phe Leu Leu Arg Asn Leu Ala Val Leu Asp Leu Cys Phe Ser Ser
 65 70 75
 Val Thr Ala Pro Lys Met Leu Val Asp Leu Ser Glu Lys Lys
 80 85 90
 Thr Ile Ser Tyr Gln Gly Cys Met Gly Gln Ile Phe Phe Phe His
 95 100 105
 Phe Leu Gly Gly Ala Met Val Phe Phe Leu Ser Val Met Ala Phe
 110 115 120
 Asp Arg Leu Ile Ala Ile Ser Arg Pro Leu Arg Tyr Val Thr Val
 125 130 135
 Met Asn Thr Gln Leu Trp Val Gly Leu Val Val Ala Thr Trp Val
 140 145 150
 Gly Gly Phe Val His Ser Ile Val Gln Leu Ala Leu Met Leu Pro
 155 160 165
 Leu Pro Phe Cys Gly Pro Asn Ile Leu Asp Asn Phe Tyr Cys Asp
 170 175 180
 Val Pro Gln Val Leu Arg Leu Ala Cys Thr Asp Thr Ser Leu Leu
 185 190 195
 Glu Phe Leu Lys Ile Ser Asn Ser Gly Leu Leu Asp Val Val Trp
 200 205 210
 Phe Phe Leu Leu Leu Met Ser Tyr Leu Phe Ile Leu Val Met Leu
 215 220 225
 Arg Ser His Pro Gly Glu Ala Arg Arg Lys Ala Ala Ser Thr Cys
 230 235 240
 Thr Thr His Ile Ile Val Val Ser Met Ile Phe Val Pro Ser Ile
 245 250 255
 Tyr Leu Tyr Ala Arg Pro Phe Thr Pro Phe Pro Met Asp Lys Leu
 260 265 270
 Val Ser Ile Gly His Thr Val Met Thr Pro Met Leu Asn Pro Met
 275 280 285
 Ile Tyr Thr Leu Arg Asn Gln Asp Met Gln Ala Ala Val Arg Arg
 290 295 300
 Leu Gly Arg His Arg Leu Val
 305

<210> 37
 <211> 314
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <225> Incyte ID No: 7472012CD1

<400> 37

WO 01/42288

PCT/US00/33382

```

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

```

```

<210> 38
<211> 310
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyta ID No: 7472014CD1

```

```

<400> 38
Met Gly Arg Asn Asn Leu Thr Arg Pro Ser Glu Phe Ile Leu Leu
1 5 10 15
Gly Leu Ser Ser Arg Pro Glu Asp Gln Lys Pro Leu Phe Ala Val
20 25 30
Phe Leu Pro Ile Tyr Leu Ile Thr Val Ile Gly Asn Leu Leu Ile
35 40 45
Ile Leu Ala Ile Arg Ser Asp Thr Arg Leu Gln Thr Pro Met Tyr
50 55 60
Phe Phe Leu Ser Ile Leu Ser Phe Val Asp Ile Cys Tyr Val Thr
65 70 75
Val Ile Ile Pro Lys Met Ile Val Asn Phe Leu Ser Glu Thr Lys
80 85 90
Thr Ile Ser Tyr Gly Glu Cys Leu Thr Gln Met Tyr Phe Phe Leu
95 100 105

```

WO 01/42288

PCT/US00/33382

```

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

```

```

<210> 39
<211> 359
<212> PRT
<213> Homo sapiens
<220>
<221> misc_feature
<223> Incyte ID No: 747202UCD1

```

```

<400> 39
Met Phe Lys Ala Ile Leu Gly His Val Trp Pro Lys Asp His Gly
1 5 10 15
Leu Asp Lys Leu Val Val Arg Cys Pro Arg His Thr Glu Pro Trp
20 25 30
Asn Leu Thr Gly Ile Ser Glu Phe Leu Leu Leu Gly Leu Ser Glu
35 40 45
Asp Pro Glu Leu Gln Pro Val Leu Pro Gly Leu Ser Leu Ser Met
50 55 60
Tyr Leu Val Thr Val Leu Arg Asn Leu Leu Ile Ile Leu Ala Val
65 70 75
Ser Ser Asp Ser His Leu His Thr Pro Met Cys Phe Phe Leu Ser
80 85 90
Asn Leu Cys Trp Ala Asp Ile Gly Phe Thr Ser Ala Met Val Pro
95 100 105
Lys Met Ile Val Asp Met Gln Ser His Ser Arg Val Ile Ser Tyr
110 115 120
Ala Gly Cys Leu Thr Gln Met Ser Phe Phe Val Leu Phe Ala Cys
125 130 135
Ile Glu Asp Met Leu Leu Thr Val Met Ala Tyr Asp Arg Phe Val
140 145 150
Ala Ile Cys His Pro Leu His Tyr Pro Val Ile Met Asn Pro His
155 160 165
Leu Gly Val Phe Leu Val Leu Val Ser Phe Phe Leu Ser Leu Leu
170 175 180
Asp Ser Gln Leu His Ser Trp Ile Val Leu Gln Phe Thr Phe Phe
185 190 195
Lys Asn Val Glu Ile Ser Asn Phe Val Cys Asp Pro Ser Gln Leu
200 205 210

```

WO 01/42288 PCT/US00/33382

Leu Asn Leu Ala Cys Ser Asp Ser Val Ile Asn Ser Ile Phe Ile
 215 220 225
 Tyr Leu Asp Ser Ile Met Phe Gly Phe Leu Pro Ile Ser Gly Ile
 230 235 240
 Leu Leu Ser Tyr Ala Asn Asn Val Pro Ser Ile Leu Arg Ile Ser
 245 250 255
 Ser Ser Asp Arg Lys Ser Lys Ala Phe Ser Thr Cys Gly Ser His
 260 265 270
 Leu Ala Val Val Cys Leu Phe Tyr Gly Thr Gly Ile Gly Val Tyr
 275 280 285
 Leu Thr Ser Ala Val Ser Pro Pro Pro Arg Asn Gly Val Val Ala
 290 295 300
 Ser Val Met Tyr Ala Val Val Thr Pro Met Leu Asn Pro Phe Ile
 305 310 315
 Tyr Ser Leu Arg Asn Arg Asp Ile Gln Ser Ala Leu Trp Arg Leu
 320 325 330
 Arg Ser Arg Thr Val Glu Ser His Asp Leu Leu Ser Gln Asp Leu
 335 340 345
 Leu His Pro Phe Ser Cys Val Gly Glu Lys Gly Gln Pro His
 350 355

<210> 40
 <211> 336
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 104941CB1

<400> 40
 atggagataa ggaactacag cagcagcacc tcaggettea tcctctggg cctctctcc 60
 aacctcagc tgcagaaacc tctctctgcc atctctctca tcalgtacct gctcctgag 120
 ggggggactg tgcctctat cccggccatc tactctggac ccaggctcca gacctctatg 180
 tactctcttc tcagcaactt gctctctatg gatactctgt tacaacacgt catagtctct 240
 agatctctgg tgaattttct atcagagaca aaggttatct cctatgtggg ctgctctgac 300
 cagatctact tctttatggc atctgggaac acctgcagct acctctgagc cctctctgac 360
 atcgaccggc tggctggcat ctgcaacccc ttacactatg atgtgtttal gaaacccagg 420
 aattgctctc tcatgctatt gggttcttgc agcatctccc acctacatc cctgtctcgc 480
 gtgctactta tctctctctt gtctctctgt gcctctcaca tcattaagca cttttctct 540
 gcaaccacgc ctgtctctaa gctctctctg tctgacacat cctccagcca gatgtgtgtg 600
 atgactggga cttctctctt cctctctctg cctctctctt gatactctt cctctctctg 660
 ggaatctctg tcaactgtgt cagaatcccc tctgcaagct ggaagtggaa gctctctct 720
 acctgagctt cccacctctc tgcagttagc cttctctctt ggaagtctat tctgtctct 780
 ttttagcccc tctctctctt cctctctctt agggaccggg taaccacagt tatgtctaca 840
 gtatgtctac cctctctctt cctctctctt taccactgga ggaacaaaga tatgaagag 900
 ggtttgaaga aattcaggga cagaatttacc cggtaa 936

<210> 41
 <211> 3365
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1499408CB1

<400> 41
 atggaccagc cagaggcccc ctgctctcagc acggggccgc gctctcgggt ggcccgcgag 60
 ctctctctgg ctgctctgga ggaactgagc caagagcagc tgaaggctt ccgcccacag 120
 ctgctctcagc tggcccggga cggagcagc atccccggg ggccgtgga ggcctcggac 180
 gccctggacc tctctggaga gctggcccag tctctcggcc cggagcctgc cctggagctg 240
 gccctcaaga cctcaagag gggggagcgc cctgactgag cggcagctt ccaggagcgg 300
 cggctcagc ggtctgctt cggctcagc acgtctctct cctgtctga ctcaagagag 360
 aggtcccggg ggtctgctt ggaactgag cctctctctt agggagga cgcctctct 420
 gtgaagatca ccaagcctt caccagctg ctctctctgc ccgagagcgc cgcctcggag 480
 gaggcctcgg ggcccggga agagcctgag cggggcgcg cgcctcctc ggacacgac 540

WO 01/42288

PC/T/US00/33382

```

accttaacc gctcttccg cdcgacagag gacggccgce ggcggctgac cgtggttctg 600
caggccccgg cgggcatcgg caagaccata gccggcaaaa agatccctga cgaactgggag 660
gagggaagc tylaaccagg ccagglygac ttccgcttct tcatgccccty gggcagctg 720
ctggagaggg cgggcacggg cagcctggct gaectgatic tggaccagtg ccccgaccgc 780
ggcctccggg tgcgcagat gctggcccaz ccccaaggcc tgccttccat cctggaccgc 840
gcgagagagc tgcggagctt gggggggccc gaggccgccc cctgcaaca cctctcag 900
ggcgagagcg gcgcgggggt gclaggcgga ctgctgagca agcagctgct gcccaagccc 960
ctccctgctg tgaaccagcg cgcgcgcgcc cccgggagge tgcaggcccg cctgtgttcc 1020
ccgcagtgcg ccgaggtcgc cgccttctcc gacaaqgaca agaaagagta ttctacaag 1080
tctctccggg atgagaggag gcccgagcgc gectaccgct tctgtaagga gaacgagagc 1140
ctgtctgcgc tgtgcttctg gccctctgtg cgtgggctcg tgtgaccgt gctgcgccag 1200
cagctggagc tgggtcggga cctgtctgac acgtccaaga ccaccagtc agtgtacctg 1260
cttttcaaca ccgggcttct gactctggct ccgctagcgg ccggcccccg gctgcaagga 1320
gaactgycra abctgtgccc cctggccagc gaggggctcc tgggaaccag gggcaattt 1380
gcggagaagc aactggagca actggagctt cgtggctcca aadtgcagac gctgtttctc 1440
agcaaaaagg agctgcgggg cgtgctggag acagaggcca cctaccagtt cctgaccag 1500
ajcttccagg agttccctgc ggcactgtcc taactgtcgg aggaccgagc ggtgccccag 1560
accgctggc gggcgttgg gacactctg cgtggggcag cccagccgca cagccacty 1620
glctcaaca ccgacttccc ctctggactg ctgagccgag agcggalgcg cgaractcag 1680
cgcactctcg cctgcatggt ctcaagact gtaagccagc aggccttgcg gctggggcag 1740
gaaacaggac agggctgccc ccagatgga ccagggggc caaaggctc 1800
ggagcaaccc auagaccaga agaggagagc gaggagagc agcccaacta cccactggag 1860
tctgtact cctctacga gacagggag gacgcttcty tgcgccaagc cctgtgcccg 1920
tcccggagc tggcctgca gcaaglycgc ttctgcgca tggacclygc tglctgagc 1980
taactgctga ggtgctgccc tctggagag gactggccc tgatcagct cagattggtt 2040
gactgcagc agaaagagaa gaagagcctn ggaagccgc tcaaggccag cctgggtggc 2100
ggcagttctc aagccaccac aaaaaccact ccagctccc tctctatcc actcttccag 2160
caactgactc cctactgccc cactctggcc agactgagc tctccactg caactctcc 2220
gaccggtctc ccgagacccl tctgagccc ctgagggcag ccccgccact gacggagctc 2280
ggcctctccc caaacagct cagtggggc ggaactgcta tgcctgagtg gggcctagcc 2340
tggccagct ccaggggca gacgctcagg gtaactgct ctgaccacca gcpaggtc 2400
cagtaccctg tgggtatgct tggcagagc cccgcccgc caacctgga tctcagggc 2460
tgcacactgc ccgccccat ggtgaactac ctgtgtcag tctgcagca caaggatgc 2520
ggctgcaga cctcagctc ggcctctgag gactgagag agcagctact acagagact 2580
caggtgtgca agagaccac ggggactg gtaactcac accagcctg ggaaggccac 2640
cccaactctc caaagaaact cctctgaaec ttctgagct ctggtggca gaggagatg 2700
aaagacccta gtaaaagctc ctgtggagag aacggcccat tccaaaggca gggagatt 2760
gctctcggcc tttgggaaac tlltgagccg agaggccga gaaaggcatg tggaggccu 2820
agacaaggca cctgcccgg tccagggcag gccagggacc gcccctctc tccaaactg 2880
ggylaccctc tctcccag ccccaact acaccaacca cctctctctc ctgagacct 2940
cagccactc ccttgaaaa cacccccca ccccaagca caataalyac agcgagact 3000
craatlaact aagcaactac ctggggcag aaataactt caetgctga tcccactg 3060
caggtggcc caagccccc ccgaactat ccaacataga ctggaggtag caggttacc 3120
gtccctccc gttaggaatg agaccatccc tggagctatg gccaccgcc acagggctc 3180
agttctgag atctltggga agggagacta gggcaggtg agacagcga gaacccccg 3240
gctgggtggg aagcatgacc acactggtgg tggagagccc caatgactg acggtaaatt 3300
ccctgtgga ctatttctg tlggtttcta tlaacctgg ccaggcctg tacaatcac 3360
gtcga 3365

```

```

<210> 42
<211> 1325
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<222> Incyte ID No. 3168839CB1

```

```

<400> 42
catggactcc ccagcctage tcccactccc actttggcac gatgttagcc aacagctcct 60
caaccaaacg ttctgttctc ccgtgtccty actaccgacc taccaccgag ctgcaacttg 120
tggctacagc ctgtgtctgt gctgcggggc tcccccaaaa cgcctagacc ctctgggtct 180
tcttgccgag gctggccgltg cactgggtgg tggagctgta caigtgtaac ctggcggcca 240
ggcaactgct ctccaccctc tggctgcccg tctgtctctc ctactacgca ctgcaacct 300
ggccctccc gnaactctct tgcggagca cggggcccat ctccagatg acaatgtag 360
gcagctgcat ctctctgatg ctcatcaag tgaaccgcta cgcgccatc gtaaccgca 420
tggcaatgag cnaactgccc aggcaccgag tggcgggctt gctctgcty ggcgtgtggg 480

```

WO 01/42288

PC/T/US00/33382

```

cgttcatacct ggtgttttgc gtgcccgcg ccgcgctgca caggccctcg cgttgccgt 540
accgggaacct cgaggtgccc ctatgcttcg agagcttcag cgacgagctg tggaaaggca 600
gactgcggccc cctogtgcctg ctgcccggg ccgtgggctt cctgctccc cttggggggg 660
tggttactacg gcccggccga gtcttctgaa cgttggcggc cccgagccc accgagagcc 720
agccggggcgg gaagaccctg cgcctctgc tggctaacct ggcattctc ctgctctgct 780
tcggggccca cagagcagg ctggcgctct agggcctctc gcaagacag ctggtgagg 840
ccagcgtgccc tgcoccgcat cggctggcgg gggtgctgat ggtgatggtg ctgctggccc 900
gcgcacaactg cgtgctggac ccgctggtgt actactttag cgcagagggc tlcggcaaca 960
ccctgcccggg cctgggcaul ccgaccggg ccaggacctc ggcaccacac gggagcggg 1020
cyygcctcgg gcaatccgaa aggtccggcg tcaccaccga cgcaccagg ccgcatgccc 1080
ccatgtcccc aggatcccg cctctgaaca cactgcaat tgcctgctc gfgcccgact 1140
ccccacgctt ctgctctgga gaggcttaca gggctctcac acagaaggt gggctgggca 1200
cttggaacct tggctggaaa ctccagctta gcaacggcga agagtcaaa ggtggagagc 1260
caggggccag ggaaggcagl gctgctggaa atgctctctt taactgtga gcaagagag 1320
cacc 1325

```

```

<210> 43
<211> 2124
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 3291235B1

```

```

<400> 43
gacctaacct ctactagtat tagaatggca atcaaatcgc gaaccgtttt aatcaaat 60
cacacatgagt ttgcaagggg acattctgag tctagcctcc gccatctccc atccccagla 120
cctaaaggtct caaagcctgt gcccagagag gaagtcacty aytgccact gccaccccc 180
cacatattct cgtgylcccc agcalcttct agcccaccag agtggagctt gctgctccc 240
gagacctggc tcaagggagg atgccacagc cgcctgcccag ctccggtctg caccatgagt 300
gatgagcggc ggtgcccctg cagtgcagtg ggcctggctg tatgtggggg cctctcccgt 360
ctggccaalg cctggggcat cctcagcgtt ggcgccaagg ayaagaaagt gaagccctg 420
gagttctctg tggctacgct cggggcacc cacatgctaa atgtggccct gccatctgct 480
accactctcc tggctcagag ggcggcctg cgcctccact tggatggaaa tggggctct 540
tgcaggtctc tctgtctccc ctctcaacc ctccactgg caactgttt cctgtctacc 600
tcccctctct accaccgcat tggatggtc tgcctgctct tcaactaccg gctgagcaat 660
gccaaagaagc agggcgtgca caagctcatg ggtatctgga tgggtctctt catctgtctg 720
gccctgctct cgttggctg gcaagcacc agcagagcgt tctacacca tggctgccc 780
tctactgctg ctgagatcgg cctggcctt ggctctgct tctgtgctct ggtggggggc 840
aacgtggcca tgggctgagt ctgcacagcc atgcctctt ccagagctt ggcctggag 900
gtgggggccc agggcggaaa ccggcctct acgtggcaca ccactgtgt ggaagaggg 960
aagggcaagc ggcctctctc catcatggc tggggcggc caaacctct tctgagacc 1020
acggcctctc tgaccaccat agtctctctc taccagctgc tcatggctt cctgtgctg 1080
gtggtgagct tcaagagcct gggggcggac gcctcagcgc cctggatggc actctgctg 1140
ctgtgtgctt cctgggccc gccctctctg ctgctctgtt tctctaggg ctcgacccc 1200
tcccggctg cctcaaaagc tgtccgggag aagtgcctg cctctatgg caacgagag 1260
gagtcagacc atgagaccag cctggaaggt ggcattctcc cggaccctgt gtlggagcc 1320
tcccggacct atggtctatg aggtactctt gtygcctag ataggatgca caatctggg 1380
atctccggcc tggaggggg cctggcccag ctctaccac tgggcccctt gcaagagac 1440
aagatgcaat acctgcaggt cccgcccagc cggcctctt cccagagca tgggagctg 1500
tgggctggcc tcccgtgccc cgcctctctg ccgcctggg gctccggcaa ggaectgtc 1560
gccctggccc acctggtgct gctgcccgg ccagagaggg ccagcggcag cctctgccc 1620
tccagggagg acyaccact gtcggcgcc cgcggccctt cggcggagag cctctgacc 1680
ctgcccctct cggcggctga tagggcccgg gggggagccc ggcactgnc cccggcagc 1740
cccgcggccc cgcggggccc cggcccggc tccgctctgg cctcctgct gcccaagcc 1800
ttggcccaga ccgctctgga gttggagcga caggccctgc gccgcccgc cggcctctc 1860
ccctctgccc cggcggccc cgaagggcga qatccggag aggcctggc gcccccanc 1920
aacgcccagc ggaagccagg gccacgccc tctgctact cgcagccgg ctctctgac 1980
cccggctgca gctgctctg gggagagccc ggggggctg ggcgggggg cgggggggg 2040
agcaccagca gcttctctggt ttcccctctc gactcctcgg gctagccac gctgactct 2100
gactctctg gctcctctc ctag 2124

```

```

<410> 44
<411> 942
<412> DNA
<413> Homo sapiens

```

WO 01/42288

PC/T/US00/33382

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472001CB1

<400> 44
atggaagaaga tcaacagcac actgttgaact gcgtttatcc tgaacaggaat tccgtatoca 60
ctcaggctaa ggcacactctt ttttgtgttc ttttttttaa tctacatcct gactcaagcty 120
ggaaacctgc ttattttaat cactgtntgg gcagacccaa ggtccatgc ccgcccocatg 180
tacatcttcc ttggtgttct ctacgtcatt gatattgagca tctctccat cattgtccct 240
cgctccatga tgaactctac tttaggtgtc aaacccatcc catttggtag ctgtgttgc 300
caactctatt tctatcactt cctggggcgc acccangtct tctctacac cctaatggcc 360
tatgcaagyt acctggcaat atgtcagccc ctggctacc ctgtgtccat gactgctaa 420
ctgagcgcct tgttggggc tggagctggc atggaagat caatccatg gctctctaa 480
gcacatccaa cctcccgctt gctctactat gggcccattc agtgagatta cttctctctg 540
gacatccctg cagltgttgg actggcctgt gctgacacaa cagtcaacga cctgtgtgag 600
ttttagacaa ttgggttggg gcttgcagct tctctctccc tgatcctct cctcctacata 660
cagatcattc aggcctctct gagaatccac acagctgatg ggcggcggc ggtcttttca 720
acttggtagg ccaatgtaac cgtgtgcaac gtgtactatg tgcctgtgc cttatctac 780
ctgagggccty aaacaacag ccccttggat gggcagctg ccttagtccc cagggcctc 840
actctctccc tcaacccct tctctcact ctggggaac aagatgtga cctggccty 900
aaagagctgc tcagaagccc aagaacctcg agtgaggttt ga 942

<210> 45
<211> 1197
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472003CB1

<400> 45
atgcacacccg tgggtacgtc cggacccaac cgtctctggg gggcacccgc caacgcctcc 60
ggctgcccgg cctgtggcgc caacgcctcc gaggccctag tccctctccc gctggccty 120
gagcctagc tcttctggc ctctctctgg gctgtgatc tcttggct gctgggaac 180
tcgtggtca tcaagtcct ctgcgcac ccagcctatg ggaacctgac caactctac 240
atgcacaccc tggcgcac ccagctgacc tctctctctg cctctctccc cttcaccgcc 300
ctcgtgtacc cctgcccgc ctggtgtctg ggcacttca tctgcaagt ctgcaactac 360
atccagcagg tctcgtgca ggcacgtgt gcaactctga ccgcatgag tctgagcgc 420
tggtaactga cgggtctccc gttgcgcgcc ctgcaccgc gcccccgc cctggccty 480
gctgtacgc tcagcactg gacagctct gggcgtgt ctgcccctg cctgccccty 540
caagcctct caagcagc cccgcctac tgcagtag ctcttccc cctggccty 600
gagcgcct tgcactga caactctg tgcctgtacc tctgcccct gctgcccac 660
tgcctctct atgcgcact cctgcccgc ctggcccgc tggcctgct ccccccgc 720
gcagatagc cctgcaagg ccaggtctg gcagcgcgc cagcgcct gggggcag 780
gtctcggcgc tgttggcgc cgtgtctct ctcttggcc cctgtgggc cccatccag 840
ctgttcttgg tctgcaagg gctggcccc ggggtctct ggcacccag cagctacgc 900
gccacgcgc ttaagcctg ggtctctct atgtctaca gaactccc cctgccccty 960
ctgtctacc cctctctgg ctgcacttc ccagagcct tcccccct ctgccccty 1020
gagcgcct gcccccgc ccccccgc ccccccgc cggacccc cggcccacac 1080
gagagctgc tccctctgg gttcccacc gcccgcaca gggcgcag gcccagagc 1140
agtgggctgg ccccccgc gctgtgctc ctggggagg acaacgcctc tctctga 1197

<210> 46
<211> 1110
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472004CB1

<400> 46
atggcccaca ctggtttgag ttccctgacc gtsalagta cagctgtgac cacacacca 60
gagactttaa agagcttaa ctggcctct cagatccc ttctgtct aatgcttct 120
atctgtttt tctcttct ttggaactt gttcttgc tcatgttta ccaaaagct 180
cgaatgagyt ctgcaattaa catctctct gccagctag ntftgaga catgtgtct 240

```

WO 01/42288

PC/T/US00/33382

```

gcagtgctga acatgccctt tgcctgyla actattctta ctaccagatg gatttttggg 300
aaattctctt glagggatc tctatgttt ttctggttat ttgtgataga gggagtggcc 360
atcctgtctc tcaattagca agataggttc ctctattatg tccagaggca ggataagcta 420
aacccatata gaggcaaggc tccgatfcca gttctctggg caacttctcl ttggttagct 480
ttctctllag ccgtagaaaa ccccgactg cagatacctt cccagagctc ccagtgtgtg 540
tttgggtacc caaccaatcc aggcctaccg gcttatgtga ttttgatttc tctatctct 600
ttctctatcc ccttctctgt aatctgttac tcaattatgg gcaatacaca caccctctgg 660
caacaatgct tyaggatcca tagctaccct gaaggtatat gctcagcca gcccagcaaa 720
ctgggtctca tgagctctga gagacccttc cagatgagca ttgacttggg cttcaaaaa 780
cgtgccttca ccaatatttt gattctcttt gctgtcttca ttgtctgtgg ggcctcattc 840
accacttacc gctctgttgg aaccatcagb aagcactttt actatcagca caacttttll 900
gagattagca cctgtctatc ctggctctgc taccctcagb ctgcattgaa tctgtgale 960
taccattgga ggaataagaa atctcaatgat gcttgccctg acatgattgc taagtctctc 1020
aagtttttgc ccagctctcc ttgttaacaa aagcagagga taactcttag tctctctat 1080
gtgtgtgggg aacatcggac gctgtgtgca

```

<210> 47
 <211> 582
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7475687CT1

```

<400> 47
nugcctatg acagylacc tgcnaatgt cagccctgc gctacccagt gctcatgaat 60
gggaggttat gcaactcct ttgtgctgga gcttgggtgc cccagctccat gcaatgctct 120
atccagacca cctgtacctt ccgcttgcct tectgtgggc ccaatcaggt agattacttt 180
atctgtgaca tcccgcagc attgagactg gctgtgtgtg acacaactgt caatgagctt 240
gtgacctttg tggacatagg gctagtgccc gccagtgtct tcatgttaat tctgtctctg 300
tatgccaaca tagtaaatgc catctggaag atacgacca ctgatgggag ggcctgggccc 360
ttctccacct gtggctccca cctaatctgt gtcacgtct actatgtccc ctglattttc 420
atctaccfta gggctggctc caaagctcc ctggctgggg ccagcgtgtg gttttacat 480
gttgtcactc cttatagaa cccctcact tatacactga ggaaccagga agtgaagct 540
gctctgaaga gataaccgc aggcacaggg gatgtaaaa ac

```

<210> 48
 <211> 519
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7483029CT1

```

<400> 48
atgtacctgg tcaactgctt cgggaacchj ctctatctcc tggccaact ctcagactcc 60
aacctccmha ccccacgtha ctctctctct tccaaactgt cctttcgaga catctgttt 120
gtgtctacca ctgtcccaaa gatctgtgtg auctatccga cacagagcag agtcaaac 180
tatgcagact gcatcaccpa gatgtgtctt ttatctctct ttgtatgttt ggacagetta 240
ctctgactgt tyatggccta tgaccggttt gtggccatct gtcacccctt gcaclacaca 300
gtcattatga actcctggct ctgtggactg ctggttctgt tctcctggat cgtgagactc 360
ctatattctc tgttacaagg nataatgcca ttgcagctgt ccttctgtac agaattgaaa 420
atccccatb ttttctgtga acttaactag gtcctccacc ttgctctctc cgaacattll 480
atcaatgaca tgatgatgaa ttttacaagt gtgtgtgtg

```

<210> 49
 <211> 663
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7477933CT1

<400> 49

WO 01/42288

PC/T/US00/33382

```

ccccgtact tctccctctc caacctgtgc tgggctgaca tgggtctcac ctcggccaag 60
gttcccaagg tgaactctga tatgcagctc catagcagag ccatctctca tgtgggcgc 120
ctgacacaga tgtcltcttt ggtccttttt geatgtatag aagycatgct cctgactgtg 180
atggccratg gctgctttgt agccalcctgt cgcctctcgc actaccagat cataglgaa 240
cctccactct gctcctctct cgtcttggta tccctttctc ttaacctggt ggafrccag 300
ctgcaaggtt ggaattggtt acaattcaac atcaaccaag atgtgaaat ctcaaatctt 360
ttctgtgacc cctctcagct tttcaacctt gcctgttctg acagctcat caatagcata 420
ttcataatt tcatagtagc kargtrtgyt ttctctcca ttccagggat ccttttctct 480
tactataaaa ttgtccctct cactctaaag atgtcatcgt cagatggaaa gtatcaagcc 540
ttctccact atggctctca cctaggagtt gtttctggt ttatlygaa agtcatlygc 600
atgtactctg cttcagcctt gtaaccaccc cccaggaaat gctgtgtggc atcagtgatg 660
tag 683

```

```

<210> 50
<211> 911
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<222> Incyte ID No: 7475164CT1

```

```

<400> 50
ggycctagtt tctcctggca ggcctgacac aacgcceaga actlcaacty ccactcttcc 60
tctgttctct tggaaatbat gtggtacacag tggtygggaa cctgggcaatg atctctctaa 120
ttgctctcag tctcaacttt lacctctcag tgtatttatt tctcagtcac ttgtcttcca 180
ttgatctctg ctactcctct gtcactacc ctaagatgct ggtgaactct gttccagagg 240
agaaacttat cctctctctg gaatcactca ctcaacttca ttctctctct attttctaa 300
ttgcagagag ctactctctg acagccatgg aatatgacc ttatgttctc atctgtctcc 360
cactgcttta caatattgtc atgtccacca gggctctctc cacaatgag gctgtgttat 420
actcactggg lllctctgga gccacaatcc atactaccgc catgtcagtg ttgtctctct 480
gtaggtctca taaggctcag cactattttt gttgatctct cccctatttg actctgtctt 540
gtccacagac ccacatcaat gagattctgc tgttccattt tggagagatt aataccttag 600
caactaacct ggcggtcctt atctctctat ctttcccttt ctctagcttc cttggtatcc 660
atccactgga gggcaatcc aaagccttgg gacctgttag ctccactctc ttgctctgag 720
gcactctttt tgggtctata acattcatgt atttcaagcc cctctccagc actactatgg 780
aaaaagagaa ngtgtctct gtgtctctaa tcaacaataa cccatgctg aactctctaa 840
tctatagctt gagyaacaag gatgtgaaa atgcactgaa gaagatgact aggggagagc 900
agtcatcctg a 911

```

```

<210> 51
<211> 932
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<222> Incyte ID No: 7473909CT1

```

```

<400> 51
caggcccccag gaaagccagt ggcctgtgta ccatgatctg atttgccttc actgtctctc 60
ctttacatccg cactctggcc acagtggttc agatcccttc agcagccagc ccccggaag 120
ccttctccac ctgtctctcc caactgggca tgggtctctc gttctatggc accggcagct 180
ccacctacat gggaccaccc acccgtact ccccgctgga agggccttg gctgctgtct 240
tctactccat cctcalacc accctgaatc cgtctactca cagcctgagc aaccagtaac 300
tgaagagagc nctgttgaag ctccatctcc ag 332

```

```

<210> 52
<211> 538
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<222> Incyte ID No: 7475252CT1

```

<400> 53

WO 01/42288

PC/T/US00/33382

```

agagccagag aatctcag gtgtcttaga attcctgctc ctgggactcc cagatgatac 60
agaccctcag ccgctcctct ttgggctgtt ccctgccaag taactgggca tggctgctggg 120
gaaccctgctc atcattctgg ccgtcagctc tgaactccat ctccacagcc ccattgactt 180
cttccctccc aacctgtcct tggctgacac cggctttggc tctactactg tcccacaagt 240
gatttggacc atccaggctc atagtagaac cactctctac gtgggctgcc tgaactcaat 300
gtctctcttg acctttttcg cactgatgga aagctctctc ctgattgtga tggcctatga 360
ccgllcgtg gccatctgtc acccctgca ctaccaagtc atcatgacc cagcactctg 420
tggctcttta gttttggtgt ctttttttct tagccttttg gactctcagc tgcacaattt 480
gattggttta caacttaact gcttcaacga tggggaatc tctaattttt ttctgtga 538

```

```

<210> 53
<211> 279
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7927572CT1

```

```

<400> 53
tcttttagat gccacagctg acaatttggat tgccttanaa atgacctgct tcaaggatgt 60
ggaaattctt aattctctct gtgaaccttc tcaactctcc cactttgcat gttgtgacac 120
cttccacaat aacataatcc tctatttccc tgaatccata ttgtgttttc ttcccacctc 180
ggggaaacct ttctcttaag ataaatttgt ttctccatt ctgaggggttt catcatcagg 240
tgggaagtat aaagccttct ccacclatgg gtctccact 279

```

```

<210> 54
<211> 291
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7481257CT1

```

```

<400> 54
atggaggtaa ccaactttgc catgtgcctg attatagttc ttgttctctt tctttatttt 60
cttttgcact atggtttcat tctgttggct gtactcaaga tcaagtcttc agcaggaaga 120
caaaaagcat ttgggacctg ttctctccat ctctgtttgg tatccacttt ctgtgggaca 180
gttacataca tgtacataca gccaggaaac agtcccaatc agaattgagg caaactcttc 240
agtatatttt actccattgt tactcccagc ttgaaacctt taalltatac g 291

```

```

<210> 55
<211> 402
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7485790CT1

```

```

<400> 55
aggatccaga actgcagccc atctctgctg ggtgtctcct gtccatgtat ctggtcacyg 60
tgetgaggaa cctctctcct agcctggctg tcaactctga ctcccacctc cacaccccaa 120
tttctcttct cctctccaac ctgtgtctgg ctgacatcgg ttccacctcg gccacggctc 180
ccaagatgat tgttggacat cygtgcata gctggagctat ctcttatgag gactgactga 240
caagatgttc ttctctgctc ctctctgcat gcttagaaga catgctctcg actgtgatgg 300
cctatgactg ttltgtgctc atctgtgccc ctctgacta cccactcctc gtgaatcttc 360
acctctgctg ctctctagtt tctgtgtcct ttctcttagc ct 402

```

```

<210> 56
<211> 639
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature

```

WO 01/42288

PC/T/US00/33382

<223> Incyte ID No: 7482993CF1

```

<400> 56
ggatcagagt gtcctcaact ggcagcaatg gcatatgac gttacattgc aatctgcaat 60
cctttaaagg atcagttat tctgagcaag gctctatga atcaattagc agcccaatgc 120
tggctcgtga gtttcttaa ctcaatggtg catcacagtg tgcacatctg cctgcccctc 180
tgtggaacaa atcagattaa ttaacttctc tctgacatcc cccctttgct gatcttctct 240
tgtggaacaa cttctgtcaa tgaattggca ctgatatcaa ctggggtctt cattggttgg 300
actcctctcc tttgtatcgt acttctctac atttgcataa cctccaccat cttgaggatc 360
cagtcctcag agggaaagac aaaaagcclt tctacatgtg cctccaccat ggcattgttc 420
ttctctttt atggcagcgc catcttllac tatgtacggc ccatctcaac ttactcaata 480
tagaaagata ggttggtttc agtgytttac agcgtctgca ccccacgtct aaacctata 540
atttaacat tgaagataa gycatcaaa gaagctgtca aaactcagg gagcaagtgg 600
cagccaccaa tttctctttt ggatagtaaa ctacttar 639

```

<210> 57

<211> 1370

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2829053CB1

```

<400> 57
cggcaccaga aggatataac cagaatcttc cagcaacatg agggagaaaa gaagaattgg 60
gcaccaacag tgggaagaga aagggagcta gacccagaa cagattggat gagcaacaa 120
gggtcctcga agaaagaaat gaagagctgc tgcagctctc cggggctca tatgacacag 180
agaaagagag ccttaccoccc tctttccggg agccacgttc taccacagag gagacatag 240
acacagctac ctacacagctg gaggtllctc agcccaaat gaagagggtg gaggagtcca 300
tctcagcccg aaactataag aaacatctcc aggattatgg gagccccagc cagttctggg 360
agcaggagct ggagagctta cactttgtca tggagatgaa gaatgagctt attcaiyagc 420
tggcaagcgg gctgaccttc atgaaacag tgaagagaaa aaatctgata ttggaggaaa 480
aaattacagc cctgcacacg gaaatagagg acctccatgt ccgaacccgc aaccaggttg 540
tctcagcccg aaactataag aaacatctcc aggattatgg gagccccagc cagttctggg 600
tgcagctcag gctcagctc cagcagagaa agagagagct gttgtaccgg tctcttgggg 660
cgaatgcttc cctcctcttc cctctgccc ctgctcctcc cactgagctt tcttctctcg 720
caccataggg tgcagggcct gggcccacca cagccctcga agtcacagct ccttccaaag 780
ttttcttggg gaagacagca ggagcctctc agttcttttc caggaagaaa cggggttggg 840
agcagatggg agatcctcgg tgtgtgccc a gtagccctg gggccttgag tfacalggaa 900
tcaccaacag gtttttggag gccccagaaa ggtctctccc tggattggc caaggganta 960
agaaagagaa gaacttctc cctctctccc gcactctgc caactccag aagttccag 1020
gctttcccag gggcagctctg tctcaagctg cccatttgac ataaaggaga cagccctcgg 1080
tcccagcttg tcagtctctc tgcagacttg ctgacttacc aactctctct aggtgtctcc 1140
actccaccct ggcctgctsa ggcctcagct ttaccctcgc attaaaatgg tgggggact 1200
ggtcaaaagga ctcttatgct actgcagtgg cccattctag gattgtctga agccagagat 1260
aggggttggg gggagtgtgg caaaccccg caaatcagag tggggaaagt gagtgttaga 1320
gagggggtct ctgaagccc ttggagctga caggcccaag cagcctcccc 1370

```

<210> 58

<211> 1567

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3068234CB4

```

<400> 58
gagggactgc gtttaatac ggagctccga ggatgtccac ttctttcca caatgaatga 50
gtgtcactat gacaagcaca tggacttttt ttataatagg agcaacactg atactgtcga 120
tgactgaaca ggaacaaaagc ttgtattat tttgtatgll gggcagtttt tctgctgttt 180
tatttttttt lolaatlctc tgytcaatgc ggcagtgatc aaaaacagaa aattctattt 240
nccccctcac taccgttgg cttaattcag tgcctccgat tctctgcctg gaattcgcct 300
tgaattctct atgttaaca caagccagct ttcaaaact ttgactgtca accgttggtt 360
tctcgtcag gggctctctg acagtagcct gactccttcc ctacccaact tgcctgttat 420
cgccgtgag aggcacatgt caatcatgag gatccgggtc catagcaacc tgcacaaaaa 480

```

WO 01/42288

PC/T/US00/33382

```

gaggggaca ctgctcaatt tgcctgctg ggcacgcgc atttttatgg gggcggccc 540
cacactgggc tggcaattgcc tctgcaacat ctctgcctgc tcttccctgg ccccaattta 600
cagcaggaat taccttgyttt totggacagt gtccaaacct atggcctccc tcaatcaggt 660
tctgggtgac ctgcgggatt acgtgtaagt caagaggnaa acccaactct tgtctccgca 720
tacaagtggg tccatcagcc gccggaggac acccatgaag ctaaagaaa cggctgatgc 780
tgccttaggg gcctttgagg tatgctggac ccggggctcg gtagttctgc tccctgagc 840
ctgcaactcc agcagatgga gcctgcagca tctgcaaaag tggttctcgc tctctggcct 900
gctcaactcc gctggaacc ccatactcta ctctacaag gacgaggaca tctatggcac 960
catggaagag atgattctgct gctttcttca ggagaacca gagaggctc cctctgcgat 1020
cccctccaca gctctcagca gtagtgacac aggcagccag tacatagagg atagtattag 1080
ccaagggtca gctctcaata aaagtacttc ctaaactctg gatgcctctc ggcccaacca 1140
ggcctccctt gggaaaagag ctgllaagaa tgattaccct tctcaacaaa agcccaatga 1200
nagtttatt tggagctctc ataatcaact gctgatcttc ttaaaaaat tttttttat 1260
agtttaaag calgtggcag aaagggagca ctgtctaat ttaggaaag cacgaaagc 1320
ggagggcttc ggcggctccc cgttctctct atgaactgct cagagcttct gtcagtccag 1380
ctgggctctc tgggtctctg cccctttctg tagccattct clltgtatt taaaaggac 1440
ttatgaaagg ctttagacca aataaacta taatgttact ttagccacct tatatagctg 1500
ctggagagat ctatgtagtr cttctctgat gcattaaaaa tgtttagaaa tgcctcaaaa 1560
aaaaaaa

```

<210> 59
 <211> 1321
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 5029478CB1

```

<400> 59
caaacctctg cgggcctcag gccgcgggaa tgtccctga algcggcggg gcagcggggc 60
ccggcctctt gcgcagcctg gacgaagcca accgcaccgc ctctcccttc tctccgcagc 120
tcaaggcaga ccaacggctg gtgctggccc cgttggagac aaccgtgctg gtagctcatc 180
ctgagtgctc gctgctgggc aacgtgtggc cctctgctgt ggtgacgccc ctagcagccc 240
ggcggcgaac tgcctgctctg gtaactcaac tctctctgc ggaactgctc tcaatcagcg 300
ctatccctct ggtgtggcct gctgctgaga ctgagcctg gctgctgggc cctgttgcct 360
gcaactctgt cttctacatg atgacccctg ggcggcagct caccatcttc accgtggccc 420
cggcagcctt ggagcgcctg gctgtcctcd tgcactgca gcggcggctg cggggtctcg 480
ggcggcgggc gggggcagtg ctgctggcgc tcaactgggg clattcggcg gtcgcccctc 540
tgcctctctg cgtctctctc cyagtcttcc cgcacaggct cccggggccc gaccaggaaa 600
ttctgatctg caaactgatt tggcccacca ttctctgaga gatctcgtgg gatgctctct 660
tgttaacttt gaactctctg gtcaccagac tggctcctgt gatcagttac tccaaaaatt 720
taagatctcc aaaggatcca agtaagagag tcaaggaag cctggctcac tggagagccc 780
accagatccc cgtgtcccag caggacttcc ggtctctcc cactctcttc ctctctatgg 840
tctctctctt cctcctgctg agcccctcca tcaaccact cctctctctc ctgctcaga 900
aectcaagca agactcctgc atctggcctt cctctctct ctgggtggg gcttccact 960
ttctcaattc agccttaaac cccatctctt acaactgac actgtgagg atgagttgga 1020
agaaaaattt ttgctgcttc tggttcccag aaaaaggagc cattttaaca gacacatctg 1080
tcaaaaagaa tgaactgtctg atactctctg gctaatctt cttaattctt cttaagca gagttctcca 1140
cactgtgaga gctgtggcat gctttaaac aggtttaa tctagttacc tcaatcagcg 1200
gacctctgt taaagaaat gaactctgca aatagacat ccacagctc gataaactca 1260
ggggctgaca caaagtttca taatatttct cctttataa aggatttctt gccaggtgc 1320
a

```

<210> 60
 <211> 1110
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 5102576CB1

```

<400> 60
atgtctctct tgaattgctc gcatcaagca taaaggctg gtttttatct tttattttc 60
caactctctt tctctctca ggtgtcaag actcaaggag cccggaaac tcaaggtgtt 120
ctgagaattc cctctctctg gactctcaga ggaatcaaaa ctgcagcggc cctctgctll 180

```

WO 01/42288

PC/T/US00/33382

```

yctgtccctg tccctgtcca tqtatcttgt caeggtgctg aggaacctgt tcaagatcct 240
ggctgtcagc tctgactgcc ccctccacac ccccatgtac ttcttccctt ccaacctgct 300
ctggcctgac atcggtttca cctcggccal ggttcccaag atgatttggg acacgcagtc 360
gcalaycaga gtcatctctc atggcggctg cctgacaag atgtctttcc tctctcttlt 420
tgaatgtata gaaggcctgc tccctgacct gatggcctat gactgclttg tagccatctg 480
tgcctctctg caclaccag tcatctgtga tctcaacct tggctctttt tggctctgt 540
gctctctctt cttagcctgt tggatttcca gctgcacagt tggatttctg tacattaac 600
catcatcaag aatgtgaaa tctcaattf ggtctgtgac ccctctcaac tttccactct 660
tgcctgtctt gacagcttca tcaataacat attcatatct ttctgatgta ctactgtttg 720
ttttcttccc atttcaggga tttttctgtc ttactataaa attgtccctt ccatttcaag 780
gattlcatcg tcaagatggga agtataaagc cttctccacc tgtgctgttc ctctagcagt 840
tgttttctgg ttttatggaa caggcaatgy catgtacctg acttcaagctg tctcaacctc 900
ccccagaaat gctgttgggg catcaatgat gtatgtctgt gtaacccat gctgaacctt 960
ttcatctgca gcttgagaaa ccgggacata caaagtgcac tggcggagct gggcagcaga 1020
gacttcgaat ctcaatgatct gttcaactct tttttctgtg tgggtgagaa agggcaatca 1080
cattcaactct cttttctgtc aaaaaaaaaa

```

```

<210> 61
<211> 1095
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 2206534CB1

```

```

<400> 61
atcaagcca acacagcgc agagagdcgc tttctctctc tgggtttctc cgaactgctt 60
tccctgagc cggctctctt cggcttctgc ctctctgctt acctctctgc cttgaggggc 120
aacctcgggc tggctgctgt ggcgtgtgac gaccgcgcgc tgcacaagcc catgtactac 180
ttctctgccc ccctggcctt ggtgagcggc ggtctcaact ctatccttgt gccccgctg 240
ctggcaaac tgcgcgacc agcctctctg ctgcgcgcga gccactgcac gccccagctg 300
tgcctatcgc tggctctggg ttgggcgaaa tgcctctctc tggcgtgat gctctctgac 360
cgcgcgcgcg cagtgtctca ccctctctgc latgcggggc tctctctccc gccctctctt 420
gcaagcttgg ccagcctctc ctggctaac ggcctcaaca actcgttgc gcaaccgcg 480
ctctgctctg agcggcctt gtcggcctc ccctctctgc acctctctat ctgtgagctg 540
ccggcgttgc tcaagcttgc ctgcggaggc gaccgagaaa ctaccgagaa caagatgttc 600
ccgcctcgcg tggctatctt gctgtctctg ttctctctca tcttggcctc ctacggcttc 660
gtggcccgag ctctctctg catgccttcc agcggagccc gaggaggggc ggtgggcaag 720
tgtgggtccc acctgacagc cgtctgctgt ttctacggct cggccatctc cactctctg 780
cagccctcgc agcctctaaa ccaggcaagg gcccaagtct tatcctctct ctacacctgt 840
gtcaacctgt ctctcaacc gctcaactac acctcaagga ataaagagct gaaggggaca 900
gcaagagcgc tgtctggag tctgggaaa gcccaagctg gctcaatgag agttggggag 960
gggagagagt ataaagcag acccaagga tggcaatcct ccttactgag taagcttaga 1020
cttcaagctg tctattttg tatgataatc tgcgaatttt gtcctaagga gtcctatggg 1080
ggatctgttt tctct

```

```

<210> 62
<211> 1665
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 3275821CB1

```

```

<400> 62
ttctacatt cactgattct ctgaaccttc ctgtctctgc ctgtcaagta gattgtatga 60
gaactccatg aqytcataca ctcaagctc ttggcctagg ataatctctc aaaaagtgtat 120
gacatgtctg caggagagga tggtagcttg cctggagatg caacagaccg tctctcccat 180
actcgttcat tcaacccalc atgtattcac caggcaccca ctctgtctcc agcaggaact 240
tggagacccc aatatggacc taactatgaa gctgacctgg gtcctactgg ccacagacc 300
cgcaacagagc ttgatgatga ggaactctac ccccaaggtg gctgggacac ggtctctctg 360
gtggccctgc tctctctctg gctgcaagcc aatgggttga tggctgtgct ggcctgtctc 420
cagccctcgc atagagctgg cagcggcttg gctctctctc tctcagctgt ggcctctctc 480
gactctctgt tctctctctg agggccttc cagatctctg agatccggca tggggagac 540
tggcctctgg ggaactctgc ctgcctcttc tactactctc tatggggctt gctctactcc 600

```

WO 01/42288

PC/T/US00/33382

```

tccagctctc tccctgctgc cgcctcagc ctgagccgt gactgctggc gctgtgccc 660
cactggtaac ctgggcaacc cccagtcgc ctgcccctct gggctctggc cgtgtctgg 720
gtgctggcca caactctcag cgtgcccctg ctggctctcc cgaaggctgc cgtctgtgg 780
taagcaactg tcaatctgct ggaattctgt gacagccagg agctgtgctt gaggatctct 840
gaggtctctg gggctctctt gcaattctct ctgctgtctg tctgcaactt gctcaccag 900
gcacagctct ctgcaactg ccaagcgcaa cagaagcag cagctgcaag gggctctgct 960
gctgtgcccg gacacatctt gtcagctat gctgtctctg ggtctgctca ccaagctggc 1020
cagctgctct acctggcctt cctgtgggaa gctactctg gctacttger ctgggagggc 1080
ctggtctact ccgactacct gatcctactt aacagctgct taagccctct cctctgctc 1140
atgaccagtg ccgactctcg gacctctgct cgtctcctgc tctgtctctt cggggcagct 1200
ccttgcgagg agcggccggg cagcttcaag cccactgagc cacagaccca gctagattct 1260
gagggtccaa ctctcccaaa gccgctgcca gaggccactt cacagatgga tctctgtgct 1320
cagctctcag tgaactccac actccagcaa cagctggatc acagccagc caaactgaa cctcctgct 1380
aacctcaggg ccaagcccaa gctggatccc acagccagc caaactgaa cctcctgct 1440
cagccacagt cagattctgt ggcacagcca caggcagaca ctacagtcca gacctctgca 1500
cctgtgctca gttctgtgct cagctctgct gatgaagctt ccccaacccc atcctcgcct 1560
cctaccaccg gggcctctga ggaaccagcc acactctcty cctctgaaag agaaagcccc 1620
agcagcacc cgcagaggc gggccggggc gagggcccc cgtga 1665

```

<210> 63
 <211> 1659
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3744167CBI

```

<400> 63
tccctctttt ccgattagt gacgtgacag agatgtgaat ggggcaggga tgtccttga 60
tggctccaag actttagctt ctggtgcgct gtgtccagc tctgattca gttgcagcg 120
lqalggacag ttgcatgaaa gctgagactc tcaactgacag tgaaaccttc aaatgaadac 180
aatccctgct tctctgcaaa ggaactctgt aggttccccc agcttcccca cttttttct 240
gtgacctgta ggcacagaa gctgtctgct tcttaccctc ccccaggaaa cagacacactg 300
ctgagctgga agactctgc ggcacagacc acagctctcc tctgtctggc ggcagctgtg 360
gggtctctg ccaacagctt cgtggtatga agcttggcag gctgagctgc tgcacaggg 420
cgacccctgg cggcaccgct tctgtctgac ctggcctgg ccacagctgc ggtgclgctg 480
ctcaccagcc tctttgtgct ctctctgacc cggcagctct ggcctgtgg ccaggcgggc 540
tgcaagcggg tgcactactt gtcgcgctc agcatgtac ccagctgctt gctcaccggc 600
ctgtctagcc tgcagcctg cctgcgaqtc accgcctctt tcttggcctc tggctgctc 660
agcccgcccc tggcccgccc cctgctgctg gctgtctgga tggccgctct gttgctgctc 720
gtcccggccc ccgtctacc ccaactgtgg agggccggg tatgcaagct atgcaacag 780
tgcagctctc acgctgctc ccaactgagc ctggagctc taaccgcttt cgtgctctt 840
ttcggctgca tctctgctg ctacagctg acgtgtgca gctgcccgg cgcctctgtg 900
ggctctgctt ggcaccccta ccaagctgct aactctctg agcggctgc agcctgctt 1020
cacccggaag gggccttggc caagctggc ggaacccccc aggcggctg agcggcaact 1080
acggccttgg ccttctcag tctagcttc aaccggctg tctactctt caccgctgga 1140
gactgtctgc cccgggagg tcccgttct ctacagctg tctcnaag ccttggggag 1200
cccggagttc ccggcctc taqgaagg accatggc tccgaactac cctcagctg 1260
aaagtggtag ggcagggccc cggcaatgga gaaccgggg gtgggatgga gaaggcaggt 1320
cgggaatggg acccttgaca gcaagccta caactctgt ccttccctg tcccltcca 1380
cccccaacc accctccaga ggtcagttt ctgggacat tggggacct tctttgacta 1440
gagttggat ctggctgggt aggatctca taacttggg gcaggccag gctcctcaa 1500
actgagggat tatgaggtg gtgatgtcc ctcttaaga ctattgtgt cttgcaagt 1560
ggcagctacc catgtgccc cttgtttacc ttgtgcaaa tagctgta 1609

```

<210> 64
 <211> 945
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472097CBI

<400> 64

WO 01/42288

PC/T/US00/33382

```

atggggaaa actggacaat tgcagtgaa ttgtttctg tggattctc agccctgtc 60
actgagctc aggtctact gttttctct ttcttgaca ttactctgt tactttaag 120
ggcaatgtcc tcatcatctt ggtcaacata gctgactctg cactacaag tctatgtac 180
ttctctcca gaaacttgc ctctctggag atagcttca acttggcat tgtgccaag 240
atgtgggga cctgtatcat tcaagacaca accctctct tcttggatg tgcactcaag 300
atgtattct tctcttctt tggagctgt gaggctgac tctgtgcaa cagtgtatc 360
gacgctacg tggccatctg tgaccctct cactaccag tcatctagg ccatatctc 420
tgtgccagc tggcagctgc ctcttggttc tcagggtttt cagtgccac tgtgcaaac 480
acatggatt tcaatttccc ttttttggc ccaaacagg tgaacctct ctctgtgac 540
agcctctct ttabtgacc ggtctgtct gacacctct tgttgaact ggaggctct 600
acagccactg tcccatctc tctcttctt tcttctgca tctgggato ctatgtccc 660
atcctctca ctatcttcc gatgcccga gctgagggga aacatcagc attctccac 720
tgttccccc acctcttgg tgtctctct tctatagaa ctgcatctc caagtattc 780
agaccctca ccaagtgtc ttctggagc aagaagctg tgtcactct tccacagtg 840
gtgactcca ttttgaacc cactatctc agtcaagga ataaagaat gaaggtgca 900
ctgaagcgc ttabccacg gacctgggc tctcagaac tatya 945

```

```

<210> 65
<211> 1098
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472008CB1

```

```

<400> 65
atggggggt ctgttgaagc tacactgaa attccagctc agatgaaatg tcatccttca 60
agaccagta ctftaaatca attaretttc tatgtgtctg tctctcact aggaagaatg 120
catgttttag aaaccaaag ctctgctyaa attagagctg gctgagagc atgtgtaca 180
ctggtaactg aggcattctc tttagaagga aatattgtta taqltcttct gtcctttaa 240
gatccaaaac tccacatccc tatgtatttc lltctl.lcca accttctctt ggttagctc 300
tgtttgacca gnaagtgtgt tcaacagatg ttgattaact tctggggccc ggaaaagcc 360
atcagctaca ttgtctglgc cactcaactc tatgtcttct tgttggcttg ggccacgaa 420
tatgtctctc tttctgtcat ggtctggat tgttatagc cagtgctca tccactgaa 480
aaaccctgca tcatgaaacc aaactttgt ctgcagcttg ctatcttggc atggggact 540
ggcttggccc agtctctgat ccagctctct gccacctctc ggttaacctc ctgtctccag 600
rsgatgggg atgagtgtgt ttgtgaagtc ccagctctya ltcagctctc cagtaactya 660
actactaca gtgaattca gatgtctat gccagtggtg tctctctgtg gatgacctg 720
atcattatcc ttctctcttc tgggtctatt gctaaaggctg tctgagaal taagtcaact 780
gcaggacaga agaaagcatt tggcactctc atctctcanc lltcttgggt tctctctctt 840
tattgcactg tncagggct ctacttcaa ccaaaaaac actatctca tgaatgggc 900
aaa.lctca ctcttttct cactgtgga acccaactc ttaactccc ctctctctc 960
ctaaagaca aggaagctca tctttgcta aaagggcca aggracagc cgtctcagag 1020
agtccaagcc ccaagcattg gcagcttcca catgggttly gtcctgtggg tgtgcaagag 1080
acaagaacty agctttga 1098

```

```

<210> 66
<211> 954
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472013CB1

```

```

<400> 66
atgagctttg cccctaalgc ttcacactct cgggtlltlt tctctcttgg gttctcgaga 60
gctaaactct cctaacactc cctcttcttc ctgttcttgg ctatttactf gaccacata 120
ctggggaatg tgcactctgt gctgtctatc tcttgggact ccagactgca ctcacccatg 180
tattactcgc tctgtgacct ctctgtgata gacatggggc tatccacagt tacactgccc 240
cagttctctg cccatttggc ctctctctac ccaaccctc ctgctgcccg ctgcttggct 300
cagttcttlt tctctctatgc atttgggggt acagctaac bgtctcttgc tgtcttggct 360
ctgctctctc atgtgacctc ctgtgacctc ctgctctctg ctcttggcaal gaatcaccaa 420
cgttctgctc gcttactagc ctgagctggc gtgggttca tactgacac ctgttgggt 480
gtggactctg tctgtctctc ttctgtgact gggatgctg gggcaacgl taacttctc 540
cactcttltl gtgaccacag gccacttctg cagcttctt gttctgacat aattctaat 600

```

WO 01/42388

PC1/US00/33382

```

gagatggcca tatttcttga ggggtggttc cttatgctgg gccctgtgc cctcaatgta 660
ctctcttatg tccgaattgg gcccgtlatt ctacgtttgc uttcagctgc tggtcgccgc 720
cgagcagctct ccaaccgtgg afcccaccct accatggttg gtttccctca cggcaaccac 780
atttgggtact acctccagcc tcccttccag aacctccact atcaggacct ggttggctca 840
gtaatgtata ctgcattac acctctggcc aacctcttg tgaatggct ccaacataag 900
gatgtaaag gtgcaactct caggtgcttt gaatgggtga aggttagacc ctga 954

```

```

<210> 67
<211> 1098
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472015CB1

```

```

<400> 67
atggaatcat ctttctcatt tggagtgtac ottgctgtcc tggcctccct cctcattgct 60
actaaaccac tagtggctgt gctctgtctg ctgttgatcc acaagaatga tgggttcaat 120
ctctgtctca ctctgattct ggtctggctt gaaccttga ttggtgtgca catctcttgc 180
ctactcaag accagaccct cggccttctt gggccaccac agaagacctt ggcagactg 240
cggatggcat ttgtcaactt ctcgcagctt cctctgtctt tcaagctcat ctgtatcacc 300
tttgaaggtt accttgccat caagcagccc ttongctact tgaagatcat gatgtggctc 360
gtggccgggg ctgcaattgc cggctgttgg ttatgtcttt acctcattgg ctctctccca 420
ctgggaatcc ccaatgttca gcagacagcc tacaaggggc agtcagctt ctttctgtta 480
tttcaacctc acttctgtgt gacctctctc tgcgttggct tcttccagc catgctctcc 540
tttctcttga tctactgata catgctcaag atgctctca tgcacagca gcaatctga 600
aagatggaac atgcaaggag catggttggg ggttatcgat ccccaccgat tccagaccac 660
ttcaaaagtc tccgtactgt gtctgttctc attggagct ttgctctata ctggaccccc 720
ttccttatca ctggcattgt gcaggtggcc tgcacaggat gtcactctta cctagtctgt 780
gaacggtaac tgtggctgct cggcgtgggc aactccctgc tcaaccact catctatgcc 840
tattggcaga aggaggtgct actgcagctc taaccatgg cctataggat gaagaaggtg 900
ctcaactcat tctctctctt tctctctgcc aggaattgtg gccacagagc gccacagga 960
agttctctgc acatctctac tatctccacc tcagagtgtg atggctaa 1098

```

```

<210> 68
<211> 930
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472016CB1

```

```

<400> 68
atgagggaaa ataaccagtc cctctacaatg gaattcatcc tcttgggagt tactgtctag 60
caggaaacagg aagattttct ctacatctct tctctgttca tttaaccat cacatigatt 120
ggaaacctgc tcaatgtctt agccctctgc tctgtgtctc gacctcaaaa ccccatgtat 180
tttctctctg ccaactctca ctgtgtgacc atctctctct catctgtacc catctcttag 240
atgctggcra accaletctt gggcagcaaa tccatctctt ttggggagtg cctaacgca 300
atgtatttca tgatagcctt ggytaacaca gacagctata ttttggctgc aatggcatat 360
galcagactg lggcactcag ccaccuactt caclacacaa caatlatgag tcccaggtct 420
tgtatctggc ttattgtggt gctcttgggt atttggaaag caatgtctct ccccacact 480
ctgctcacag ctagtctgtc ctctctgtgc aaccaggaag tggccaactt ctactgtgac 540
attaccctct tyclgaagtt atcctgttct gacatccact ttcabygaa gatgatylac 600
ctaggggtctg gcatcttctc tggccatga ctatgcata tggctctca tatctgagtc 660
ttctcaagc tcttcaaggt tctctcaacc aagggctcct tcaagcctt ctcaactgt 720
cgttccacc tcaaggttgt ctcttcttat tatggtacag tcaatggacc gtaattctgc 780
ccttggcna attatagctt aaaagacgca gtgatcactg taatgtacac ggcagtgacc 840
ccaatgttaa atccttctat ctacagctg agaaatcggg acatgaagc tgccttgcgg 900
aaactcttca acaagayaat ctctctgtaa 930

```

```

<210> 69
<211> 711
<212> DNA
<213> Homo sapiens

```

WO 01/42288

PC/T/US00/33382

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472017CB1

<400> 69
atggagatga caaacagcag tgtaaaagga gaattcaacc tgetgetgtg gaacctaaaa 60
ggactctgac aaacaatcac attctctggg tgtgtcacc agctctacat ctccctggca 120
ttgggtccca ctgagctgtt cccctctggt gaattggctt ttgatcctga tctgtcagtt 180
tgcaaacctc tccactatac cgcctgaatg aacctccagg tctgcccagg tctgtccagtt 240
gttctgtggc tgaattggayt gggaaacaat ctatccagg gcactgacac cctctggctt 300
cctctgtgtg gacaccgatt gctcaaacat ttctctcttg catgtgagg caaccatgat 360
aatggagttc agctctctgt tgcctcactg gctctgtctc ccttgccctt agtgcataa 420
ctctctctct atggacatct agcctaaagt ggcataagga taagttaagt ccaggcttg 480
tgcaaaagcc ctgggacatg tggatcccaz ttgatagtag tctccctctt ctgtgggacc 540
atcacagctg tctacatcca gtcccaacgt tcttatgccc atgctcatgg gaagtctacc 600
tccctctctt ataccagttt gaccctgacc ctcaatctcc tcatctaac acgtggagat 660
aalgaagcga naggagcact gcattattt aacagagact taggcacata a 711

```

```

<210> 70
<211> 1092
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472018CB1

<400> 70
atgggcccgg ccgagggcct gctggccggt ctctctgtga tggtaactgg cgtggcgtg 60
ctatcccacg cactctgtgt gctttgttgc gctcaacagg ctgagctccg cactccagcc 120
tawggctcc tctctgtgaa tclgtctctg ggcacactgc tctgtggcgg cctggacatg 180
oactccagc tgcctgggtt gatctggcgg ccgacacagt cggctccagg cgnatgccaa 240
gtcatggctt tctctggcac ctctctggcg tcccaagcgg cgtgagcgt ggcggcgtg 300
agcgagacc agtggctggc agtggcttcc caactctgct agcggcagc cctggagccg 360
cgtctagcgg cctctgtgtt gggctgtgcc tgggacagt cgtggcctt ctccggcgt 420
gcactggctt gctctgtgct tggctacaag agcgcttgg cgtctgttc gctgacatg 480
ccgcccagc ctgagcttcc gctctccgca gctctcaacc ccaactccca tgcctggggc 540
tctgtctgct cgtctggcgt gctctgcttc acctgctcc aggtgaccc ggtggcagc 600
agacactgcc agcctatgga cacctgcaac atgaaggccc tccgctgtct cggcgcctg 660
caccctcgtg tggcgcagcg ctgctctacc cagcagaagg ggcgcgcaca cggcgcacc 720
aggaagcttg gaattgctat ttggaccttc ctcatctgct ttgcccgtta tctcatgacc 780
agcttggcgg agctcgtgac ctctgcaacc gtgacccc agtggcctt ctccagcaag 840
tgcctaacct acatcaagtc gctggcagac cgtctcaagt actctctgct ccgcccggc 900
ttccgccaa gctctggcgg catggtgcaz cgtctctgta agagacccc ggcgccaca 960
tccaccatg acagctctctt ggaatggccc ggcctgtgac accagctgct gaagagacc 1020
ccgcccagc cgtccaccca caacggctct gtggcacag agaatgatcc ctgctgtcag 1080
capacact ga 1092

```

```

<210> 71
<211> 927
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472019CB1

<400> 71
atggccatgg caaatgtcac agcagtgctt cagttctctc ttattggcat tctraactat 60
ctccaatgga gacacagctt ttccacatta gctgctgata ttaccctcag cacattgttg 120
gggaattggat ttatgactct tcttattcac ttgacocca acctccacc tccaatctac 180
ttctctctca gtaaccctgc tttcttagac ctctgttatg gnaacgttcc catgcccag 240
gccttggctc attgtctctc taacctctcc taacctctct atcccagatg ttggctcaa 300
accagtgctt ccttgctttt ggcacacaga gatttctctc taactgctgc catggcclat 360
gacgtgggg ttgctatcag caatcccag cgtctctcag tggctatgaa ttgcccagta 420
tgtctctctt tggctgctac ctatggggz acataactg tctcaactgc catgctacc 480
ctatctctga ggttcaact ctgtggggc aatgcaata accattttgc ctgtgagatt 540

```

WO 01/42288

PC/T/US/00/33382

```

ctctccctca ttaagctgac ctgtttgat accagcctca atgaatttat gatccctcacc 600
accagttatct tcaccctgct gctaccattt gggtttgttc tctcttctca caccagatt 650
gctatggcta tcaaaaggtat ccyctcaact cagggcaggc tcaaggcctt taccacatgt 720
ggctctcacc tgaacctggat gacaactctt tatgggtcag ccattctccat gtatatgaaa 780
actcagctca agtctctacc tgaaccagac aagtttatct cagtgtttta tggagctttg 840
acacacactg tgaacctcct gatatatagc ctgagaaaaa aagatgttaa acggycaata 900
aggaaagtta tgttgaaaag gacatgja          927

```

```

<210> 72
<211> 1032
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No. 7472021CB1

```

```

<400> 72
atgcatttcc ttoctactgt ctttggcttc ctaaacagag tcacacttgy tatcttcaga 60
ggagctatgg tcaacttgaq ttaaabjag ggtatccctc tcttgggggt tctggaatga 120
gtaagcttc agatctttac tgcactgcta ttctcggtag caccctcctt ggctcttca 180
ggcaactctc tcatatcac catcattacc gtggcccgfc gctctcattc ccccaatgat 240
tactttttca agcaactctc tctcttggac ctctctctca tctctctcac agtccccag 300
tccattgcaa attcacttat gggcaacggt taccatttcc ttgttcagtg cattctccag 360
gtttttctct tcabagctct ggctctcaea gaagtggcca tttctcaggt gatgtcttat 420
gacaggtacg cagcaactcy tcaaccactt cattalagaa ctattatgga tcccttctcc 480
tctaggcttg cagtgatagc tctgttggat gctgggggpc tctctgggct catgcttctc 540
gcaataact tctcattacc tctctgtggg aagaggtcaa ttaaccact ctctctgaa 600
gtctctcaga tgcgaaact agcctgttct batgaaftca ttaaagatag tgcacttctt 660
gcatctcaaa agtctcagc atttatctgt ttgatctca tttgtctctc ctcaattctc 720
attctctcta cagtctgag aatccccca gctgagggcc ggaaccaagt cttctccacc 780
tgcctaccar acctatttgt agccacctc tttctctcag ctgcaggttt tgaattctcc 840
agactgcctt ctgattctct atcgacttgy gaccttgtat tctccttatt ctactctgtg 900
atcccccaaa cactcaatcc agtcatctat agcttaegga atgattccat gaaggtcagca 960
ctgaggaaga tctgttcaaa ggaagagctt cctcagaaaa aaatgtgctt aaaaacctg 1020
tttaactctt ga          1032

```

```

<210> 73
<211> 972
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No. 7472009CB1

```

```

<400> 73
atgtggcaga ayaatcagac ctctctggca gaattcctac ttgaggggct cttcgtatgac 60
tccctaccct accttttctt tctctctctg accatggatg tctctctctc tctcgttga 120
ggcaaccacc tcaccattct tctctctctg attgatctca agcttcaatc accaatgta 180
ttctctccca gctagctctc cctctctggat ctgatctctc tctcccaaac catctctgaag 240
atgcttacc aactacatate tggcaagaaa tctctctctt ttgtggctgt tgcaccccag 300
caactctctct atttctgtct aggtggctca gaattctctc tctttagctct catgtctcat 360
gacgctatg ttgcatctct tcatccactg cgtctatctg tgcctalgaa caagaaggtg 420
ggactgatga tggctgtctat gtcctgttly gggcctctcg tgaactcctt aattcaatg 480
gcatcttga tgcactctcc tttctgtggg cctcggaaag tctaacactt ctactctgag 540
ttccagctg ttgtgaantt ggtatgltgc gacatctag tctatgagac caccatgac 600
ctcagagca ttctctctct cctccccatc ttctctgatt ctactctcta tgtcttccc 660
ctcaaatgt tcaatccagat gcctctctct gggagcaaga gaaatgactt tgcacttctt 720
ggctccccac tcaaggtggt ttctcttctt ttctgtctct gcatctctcc ctacatgaga 780
cncaggtccc agtgcactct attgcagaac aaagtctggt ctgtgttcta cagcatcatt 840
acgcccacat tgaattctct gatttatact ctccggaaata aagatgttagc taaggtctctg 900
anagagagtc tgaggagaaa tcttataccc cagtctalle aacgacttca attgtgttg 960
cccagatgct ag          972

```

```

<210> 74
<211> 900

```

WO 01/42288

PC/T/US00/33382

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472010CB1

<400> 74
atggaactgy aagygagactt ccttggtagt gtgggagaat tgggccaagt gatccagacc 60
tgttcgggga tctatgtgtt cactgtggtg ggnaaactgg gottgatcac cttaattggg 120
alaaaacctc gcttccacac ccccatgtac ttttccctck tcaactgttc clttatagat 180
ctctgtlact cctgtgtgtt tacccccasa atgctgaacty acttctgttc gaaaagtatc 240
actctctatg tgggatgtat gaeragata ttttctctck gtttcttgg caattctgag 300
tactatgtga ccatgtcccc aagggtctgc tttctgtga tgtttggfhe ctatgtggtt 360
tacctgttca ccatgtcccc aagggtctgc tttctgtga tgtttggfhe ctatgtggtt 420
gggtttgctg gggccatggc ccacactgga agcatgtgc gactgacctt ctgtgattcc 480
aacgtcattg accattatct ggtgtacgtt ctcacctctc tgcagctctc ctgcaaccgc 540
accatgttca gtgagctggt attttctatt gttgttggag taatcaaat gctatccagc 600
ataagcatcy tcatctctta cgttttgata ctctcaaca tctctgttat tctctctgca 660
gggggcagat ccaaacctt tagccatgg ggtcccaca taattgtgtg tgcctgtttt 720
tttgggcagc ggaattcac ctacttaaca acatctcttc ctggctctat gaacctggc 780
agatttgcct cagttcttta ccccaatgtg gttcccacy ttaaccttc gatctacgt 840
ttgaggataa aggatgataa acttgccttg gccaaccac tgaagagagt gctcttctaa 900

<210> 75
<211> 924
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472011CB1

<400> 75
atggaacaau ggaacctcac gtgggtatca gactttgtct tctctgggct ctgcysagat 60
cgggggctcc aggttttctc gtrrntaatg ttctgttttg tctactctac cactgttatg 120
ggaacatccc tlatcatcat caacgtgacc tctgatccc agctccacac acccatgtac 180
ttctctctcc gaaacctggc tctctctgac ctctgtttct ctctcagcac tgcctccaaa 240
atcttagtgg acctctcttc tgaagagaaa accatctctt accagggctg catgggtcag 300
atcttctctt tccacttttt gggagggtcc atggtctctt tctctcagc gatggccttt 360
gaeragctca ttgcaactc ccggccctcc cgttatgtca ccglnatgaa cactcagctc 420
tgaatgggac tgaatgggac caactggatg ggaagcttgc tcaactctat tctcagctg 480
tctctgagcc tccactgtcc ctctctggc ccaaacattt tggataactt ctactgtgat 540
gttcccacag tactgtgact tgcctctact gacacctcac tgcctgagtt cctcaagatc 600
tccaaagctg ggctgctgga tctctctggy ttctctctcc tctgtgatgc ctaactatc 660
atctctgtga tctgtgagtc acatccaggg gaggcaagaa ggaagccagc tccacctgc 720
accacctaca tcatctgtgt tccatgtatc ttctgtccaa gcatttacct ctatgcccgg 780
ccctcaactc catctccat ggaacagctt gttctcatcg gcccaacagt catgacctcc 840
atgtccaacc ccatgatcta laactgtagg aaccaggaca tgcagggcag agtgagaaga 900
ttagggagac accggctggt ttga

<210> 76
<211> 545
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472012CB1

<400> 76
atggacaaca gcaactggac cagtgtgtcc cattttgttc tcttgggcat ttccaccac 60
ccagaayagc aaatcccact ctctctgttt ttctcactca tctaagcaat caatalttct 120
ggcaacttgg ccatcaccac actgallctc tctgtctcac gctcccaat ccccatgtac 180
acttctctca gaaactggc ctctgaaagc atgtgcttca cctcaaccac gttcccacag 240
atctctcaga ttattttctc cctcaaaag gtaatttctt acacagctg tttagccca 300
acttatttct tcatttgctt ngcctcatg gaaaactta tctctggctg gatggctat 360

WO 01/42288

PC/T/US00/33382

```

gacaggataga ttgcaatctg ccaacccttc caactacata tgatgctgac tagaattgctg 420
tgctgaaga tgggtgctat gtgcaatcct cctctcgaac tcatgcaat gctgcaaac 480
ttctccatag gccaactaat ctctctgga gataaacgaa tcccaccctt ctctctgac 540
ctctacgctc tgatgaagat ctctctgacc agcaactaac tcaaacacct tatgattcac 600
acagaaggct ctgttgtaat cagtgaggct ctggcctca ttactggctt ctatgctgc 660
atcatctggt tggctctcgg galcacctca gccaaggga ctcaaggga ctctctacc 720
tgcggcttcc acctcactgt ggtggocata ttctatgga cctcagttg ggtctactc 780
cgccaccttt ccagctattc agtgaccaag ggtcgcata taacagtcgt gtacacagt 840
gtgactccca tgcgtaacc ctctcatctc agcctgagga atggggatgt caagggagg 900
ttcatgaaat ggatgagcag aatgcagact ttttcttta gataa 945

```

<210> 77
 <211> 933
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472014csl

```

<400> 77
atgggaagaa afaaccatac aagaccctct gaattctacc tctttgact ctctctcga 60
ctctgggctc agaagccgct ctctctctgt tctctccca tctacttat cacagtata 120
ggaaaccctc ttatcctctc ggcctctcgc tcaagacctc gtctccagac gccctgtac 180
ttctctcaa gcatcctgct ttttctgac atttgctatg tgacaglac tatccctaa 240
atgctgglya acttcttacc agagacaaag acctctctt acggtgagtg tctgaccag 300
aagtctcttt tcttgcctt tggaaacaca gacagttacc tctgacagc catggccatt 360
gaccctatg tggcaatg caatctctc cactacatca ccatctgag tcaagatgc 420
tgtgctctgc tctggtctct ctctctctgc tctctctctc tctctctctc tctctctctc 480
ctctctgact atcagctcat ctctctctgc tctctctctc tctctctctc tctctctctc 540
gataaacag tgcataaatt gtctctctc tctctctctc tcaagaact ccaagtaag 600
acagaagctc tggctgctat aatgcctcgc tttctctcga tcatctctc tctctctctc 660
atctctctca ctgtctctga gattctctca gctctctgaa agcgttaagc atttctctc 720
tctctctctc atctctcagt ggtgacctgt tttctctgaa gattctctc tctctctctc 780
cagctctctc ccaactctac tctcaagac tctcaagac caaatctctc cactctctc 840
actctctctc caactctctc tctctctctc ctgagggaca aagacatgaa gcaaggtctg 900
gcaagttga tgcacaggat gaaatgctag taa 933

```

<210> 78
 <211> 1080
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472020cbl

```

<400> 78
atgttcaag ccatcctctg ccatctgttg ccaaaagacc atgggttggg caagcttgtt 60
gttaaggtct caagaacac agagcctctg aactctcag gcatctcaga attctctctc 120
ctgagctctc cagaggatcc agaactctag cctgctctc ctgggtctgc ctgtctctc 180
tacctgctca cgttctctag caactctctc atctctctct ctgtctctc tgaactctca 240
ctcccaacct ccatctctct ctctctctct aactctctct gggctgcatc cgtctctctc 300
tggcaactgg tcccaagat gattctctga atgcagctgc atagcagct calctctctc 360
gggggtctgc tgacacagat gctctctctc gctctctctc catgtacaga agacatctc 420
ctgacagtga tggctctctg ccgattctgt gcatctctc accctctctc ctaccctctc 480
atcatgact ctcaactctg tctctctctc gtttctctct cctctctctc cagctctctc 540
gattctctct tgaagctctg gattctctct caactctct tctcaagaa tctggaactc 600
tcaattctct ctgtctctct atctctctct ctcaactctc ctgtctctca caagctctc 660
aatgactctc tcalatattc agatctctct atgttctctt tctctctctc tctgagctc 720
ctttctctct acgctaacaa tctctctctc atctcaagaa tttctctctc agatctctc 780
tctcaagctc tctcaactct tggctctctc ctgctctctc tttctctctc tctgagctc 840
ggctctctct tctcaactct tctcaactct tctcaactct tctcaactct tctgagctc 900
tcaatctctc tctcaactct tctcaactct tctcaactct tctcaactct tctgagctc 960
atgtctctct caagctctct gttgagctct cgcagctctc caagctctc tctgagctc 1020
tctcaactct atctctctc tctctctctc tgtgtctctc agaatctca accactctc 1080

```

【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
14 June 2001 (14.06.2001)

PCT

(10) International Publication Number
WO 01/42288 A3

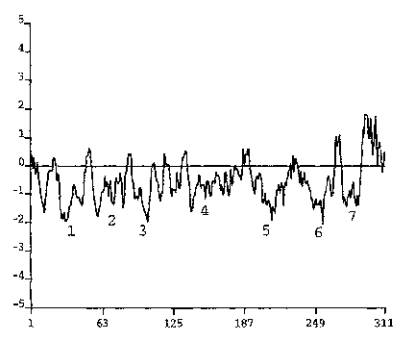
- (51) International Patent Classification: C07K 14/705, C12N 15/12
- (21) International Application Number: PCT/US00/33382
- (22) International Filing Date: 7 December 2000 (07.12.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
 - 60/172,852 10 December 1999 (10.12.1999) US
 - 60/171,732 22 December 1999 (22.12.1999) US
 - 60/176,148 14 January 2000 (14.01.2000) US
 - 60/177,533 21 January 2000 (21.01.2000) US
- (71) Applicant *for all designated States except US*: INCYTE GENOMICS, INC. (US/US); 3160 Piner Drive, Palo Alto, CA 94304 (US).
- (72) Inventors: and
- (75) Inventors/Applicants *for US only*: BURFORD, Neil
- (74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc.; 3160 Piner Drive, Palo Alto, CA 94304 (US).
- (81) Designated States (*exceptional*): AF, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CL, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, 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, TL, 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

[Continued on next page]

(54) Title: G-PROTEIN COUPLED RECEPTORS



WO 01/42288 A3



SEQ ID NO:1
(Incyte ID No. 104941CD1)

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

WO 01/42288 A3



patent (AM, AZ, BY, KG, KZ, MD, RU, TL, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IL, IT, LI, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(88) Date of publication of the international search report: 14 February 2002

Published:

— with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT		Internat. Application No. PCT/US 00/33382
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C07K14/705 C12N15/12		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 7 C07K		
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) BIOSIS, EPO-Internal, NPI Data, STRAND		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 92 17585 A (UNIV COLUMBIA) 15 October 1992 (1992-10-15) claims 37.5; figure 10 ---	1-28
X	BUCK L ET AL: "A NOVEL MULTIGENE FAMILY MAY ENCODE ODORANT RECEPTORS: A MOLECULAR BASIS FOR ODOR RECOGNITION" CELL, US, CELL PRESS, CAMBRIDGE, MA, vol. 65, 5 April 1991 (1991-04-05), pages 175-187, XP00229935 ISSN: 0092-8674 the whole document --- -/--	1-28
<input checked="" type="checkbox"/>	Further documents are listed in the continuation of box C.	<input checked="" type="checkbox"/> Patent family members are listed in annex.
* Special categories of cited documents:		
<p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" documents referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the prior art or theory underlying the invention</p> <p>"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</p> <p>"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</p> <p>"A" document member of the same patent family</p>		
Date of the actual completion of the international search 31 May 2001	Date of mailing of the international search report 23.08.01	
Name and mailing address of the ISA European Patent Office, P.O. Box 5518 Patentstr. 2 NL-2280 HV Rijswijk Tel. (+31-70) 348-2040, Te. 31 651 epo nl, Fax: (+31-70) 348-9316	Authorized officer Meyer, W	

Form PCT/ISA/210 (enclosed sheet) July 1992

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/US 00/33382

C(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MCDONALD TERRY ET AL: "Identification and cloning of an orphan G protein-coupled receptor of the glycoprotein hormone receptor subfamily." BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, vol. 247, no. 2, 18 June 1998 (1998-06-18), pages 266-270, XP000999191 ISSN: 0006-291X the whole document ---	1-28
A	GONG Z ET AL: "Rapid identification and isolation of zebrafish cDNA clones" GENE: AN INTERNATIONAL JOURNAL ON GENES AND GENOMES,GB,ELSEVIER SCIENCE PUBLISHERS, BARKING, vol. 201, no. 1-2, 12 November 1997 (1997-11-12), pages 87-98, XP004126460 ISSN: 0378-1119 abstract ---	1-28
A	STADEL J M ET AL: "Orphan G protein-coupled receptors: a neglected opportunity for pioneer drug discovery" TRENDS IN PHARMACOLOGICAL SCIENCES,GB,ELSEVIER TRENDS JOURNAL, CAMBRIDGE, vol. 18, no. 11, 1 November 1997 (1997-11-01), pages 430-437, XP004099345 ISSN: 0165-6147 abstract; table 1 ---	1-28
A	O'DONN B F ET AL: "DISCOVERY OF THREE NOVEL G-PROTEIN-COUPLED RECEPTOR GENES" GENOMICS,ACADEMIC PRESS, SAN DIEGO,US, vol. 47, no. 2, 15 January 1998 (1998-01-15), pages 310-313, XP000863706 ISSN: 0888-7543 abstract ---	1-28
A	MARCHESE A ET AL: "Novel GPCRs and their endogenous ligands: expanding the boundaries of physiology and pharmacology" TRENDS IN PHARMACOLOGICAL SCIENCES,GB,ELSEVIER TRENDS JOURNAL, CAMBRIDGE, vol. 20, no. 9, 1 September 1999 (1999-09-01), pages 370-375, XP004178194 ISSN: 0165-6147 abstract ---	1-28
	---	-/--

2

Form PCT/ISA210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT		International Application No. PCT/US 00/33382
C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	LEE DENNIS K ET AL: "Cloning and characterization of additional members of the G protein-coupled receptor family." BIOCHIMICA ET BIOPHYSICA ACTA., vol. 1490, no. 3, 29 February 2000 (2000-02-29), pages 311-323, XP000999198 ISSN: 0006-3002 the whole document	1-28
P,X	ROUQUIER SYLVIE ET AL: "The olfactory receptor gene repertoire in primates and mouse: Evidence for reduction of the functional fraction in primates." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES, vol. 97, no. 6, 14 March 2000 (2000-03-14), pages 2870-2874, XP002168634 March 14, 2000 ISSN: 0027-8424 the whole document	1-28

2

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT	In. International application No. PCT/US 00/33382
Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)	
This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:	
1. <input checked="" type="checkbox"/> Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely: Although claims 18, 21, 24 are 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.	
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.: partially 1-28	
Remark on Protest	
<input type="checkbox"/> This 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 00/33382

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

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

1. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.1, its corresponding DNA (SEQ.ID.40) and the use thereof.

2. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.2, its corresponding DNA (SEQ.ID.41) and the use thereof.

3. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.3, its corresponding DNA (SEQ.ID.42) and the use thereof.

4. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.4, its corresponding DNA (SEQ.ID.43) and the use thereof.

5. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.5, its corresponding DNA (SEQ.ID.44) and the use thereof.

6. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.6, its corresponding DNA (SEQ.ID.45) and the use thereof.

7. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.7, its corresponding DNA (SEQ.ID.46) and the use thereof.

8. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.8, its corresponding DNA (SEQ.ID.47) and the use thereof.

9. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.9, its corresponding DNA (SEQ.ID.48) and the use thereof.

International Application No. PCT/US 00/33382

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

10. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.10,
its corresponding DNA (SEQ.ID.49) and the use thereof.

11. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.11,
its corresponding DNA (SEQ.ID.50) and the use thereof.

12. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.12,
its corresponding DNA (SEQ.ID.51) and the use thereof.

13. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.13,
its corresponding DNA (SEQ.ID.52) and the use thereof.

14. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.14,
its corresponding DNA (SEQ.ID.53) and the use thereof.

15. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.15,
its corresponding DNA (SEQ.ID.54) and the use thereof.

16. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.16,
its corresponding DNA (SEQ.ID.55) and the use thereof.

17. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.17,
its corresponding DNA (SEQ.ID.56) and the use thereof.

18. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.18 its
corresponding DNA (SEQ.ID.57) and the use thereof.

International Application No. PCT/US 00/33382

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

19. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.19,
its corresponding DNA (SEQ.ID.58) and the use thereof.

20. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.20,
its corresponding DNA (SEQ.ID.59) and the use thereof.

21. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.21,
its corresponding DNA (SEQ.ID.60) and the use thereof.

22. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.22,
its corresponding DNA (SEQ.ID.61) and the use thereof.

23. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.23,
its corresponding DNA (SEQ.ID.62) and the use thereof.

24. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.24,
its corresponding DNA (SEQ.ID.63) and the use thereof.

25. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.25,
its corresponding DNA (SEQ.ID.64) and the use thereof.

26. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.26,
its corresponding DNA (SEQ.ID.65) and the use thereof.

27. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.27,
its corresponding DNA (SEQ.ID.66) and the use thereof.

International Application No. PCT/US 00/33382

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210	
28. Claims: partially 1-28	G protein-coupled receptor as characterized by SEQ.ID.28, its corresponding DNA (SEQ.ID.67) and the use thereof.
29. Claims: partially 1-28	G protein-coupled receptor as characterized by SEQ.ID.29, its corresponding DNA (SEQ.ID.68) and the use thereof.
30. Claims: partially 1-28	G protein-coupled receptor as characterized by SEQ.ID.30, its corresponding DNA (SEQ.ID.69) and the use thereof.
31. Claims: partially 1-28	G protein-coupled receptor as characterized by SEQ.ID.31, its corresponding DNA (SEQ.ID.70) and the use thereof.
32. Claims: partially 1-28	G protein-coupled receptor as characterized by SEQ.ID.32, its corresponding DNA (SEQ.ID.71) and the use thereof.
33. Claims: partially 1-28	G protein-coupled receptor as characterized by SEQ.ID.33, its corresponding DNA (SEQ.ID.72) and the use thereof.
34. Claims: partially 1-28	G protein-coupled receptor as characterized by SEQ.ID.34, its corresponding DNA (SEQ.ID.73) and the use thereof.
35. Claims: partially 1-28	G protein-coupled receptor as characterized by SEQ.ID.35, its corresponding DNA (SEQ.ID.74) and the use thereof.
36. Claims: partially 1-28	G protein-coupled receptor as characterized by SEQ.ID.36, its corresponding DNA (SEQ.ID.75) and the use thereof.
37. Claims: partially 1-28	

International Application No. PCT/US 00/33382

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

G protein-coupled receptor as characterized by SEQ.ID.37,
its corresponding DNA (SEQ.ID.76) and the use thereof.

38. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.38,
its corresponding DNA (SEQ.ID.77) and the use thereof.

39. Claims: partially 1-28

G protein-coupled receptor as characterized by SEQ.ID.39,
its corresponding DNA (SEQ.ID.78) and the use thereof.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 00/33382

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9217585 A	15-10-1992	AU 669187 B	30-05-1996
		AU 1796192 A	02-11-1992
		CA 2106847 A	06-10-1992
		EP 0578784 A	19-01-1994
		JP 6569702 T	02-11-1994

フロントページの続き

(51) Int.Cl. ⁷	F I	テーマコード(参考)
A 6 1 K 48/00	A 6 1 P 1/00	4 C 0 8 4
A 6 1 P 1/00	A 6 1 P 1/02	4 C 0 8 6
A 6 1 P 1/02	A 6 1 P 1/04	4 H 0 4 5
A 6 1 P 1/04	A 6 1 P 1/06	
A 6 1 P 1/06	A 6 1 P 1/08	
A 6 1 P 1/08	A 6 1 P 1/10	
A 6 1 P 1/10	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 1/18	
A 6 1 P 1/18	A 6 1 P 3/00	
A 6 1 P 3/00	A 6 1 P 3/04	
A 6 1 P 3/04	A 6 1 P 3/10	
A 6 1 P 3/10	A 6 1 P 5/18	
A 6 1 P 5/18	A 6 1 P 5/38	
A 6 1 P 5/38	A 6 1 P 7/00	
A 6 1 P 7/00	A 6 1 P 7/06	
A 6 1 P 7/06	A 6 1 P 7/10	
A 6 1 P 7/10	A 6 1 P 9/00	
A 6 1 P 9/00	A 6 1 P 9/04	
A 6 1 P 9/04	A 6 1 P 9/10	
A 6 1 P 9/10	A 6 1 P 9/12	
A 6 1 P 9/12	A 6 1 P 9/14	
A 6 1 P 9/14	A 6 1 P 11/00	
A 6 1 P 11/00	A 6 1 P 11/06	
A 6 1 P 11/06	A 6 1 P 13/02	
A 6 1 P 13/02	A 6 1 P 13/08	
A 6 1 P 13/08	A 6 1 P 13/10	
A 6 1 P 13/10	A 6 1 P 13/12	
A 6 1 P 13/12	A 6 1 P 15/00	
A 6 1 P 15/00	A 6 1 P 15/08	
A 6 1 P 15/08	A 6 1 P 17/00	
A 6 1 P 17/00	A 6 1 P 17/06	
A 6 1 P 17/06	A 6 1 P 17/16	
A 6 1 P 17/16	A 6 1 P 19/00	
A 6 1 P 19/00	A 6 1 P 19/02	
A 6 1 P 19/02	A 6 1 P 19/04	
A 6 1 P 19/04	A 6 1 P 19/10	
A 6 1 P 19/10	A 6 1 P 21/00	
A 6 1 P 21/00	A 6 1 P 21/04	
A 6 1 P 21/04	A 6 1 P 25/00	
A 6 1 P 25/00	A 6 1 P 25/02	
A 6 1 P 25/02	A 6 1 P 25/02	1 0 3
A 6 1 P 25/04	A 6 1 P 25/04	
A 6 1 P 25/08	A 6 1 P 25/08	
A 6 1 P 25/14	A 6 1 P 25/14	
A 6 1 P 25/16	A 6 1 P 25/16	
A 6 1 P 25/18	A 6 1 P 25/18	

A 6 1 P 25/20	A 6 1 P 25/20	
A 6 1 P 25/22	A 6 1 P 25/22	
A 6 1 P 25/28	A 6 1 P 25/28	
A 6 1 P 27/02	A 6 1 P 27/02	
A 6 1 P 29/00	A 6 1 P 29/00	
A 6 1 P 31/04	A 6 1 P 29/00	1 0 1
A 6 1 P 31/12	A 6 1 P 31/04	
A 6 1 P 31/16	A 6 1 P 31/12	
A 6 1 P 31/18	A 6 1 P 31/16	
A 6 1 P 31/22	A 6 1 P 31/18	
A 6 1 P 33/00	A 6 1 P 31/22	
A 6 1 P 33/02	A 6 1 P 33/00	
A 6 1 P 33/14	A 6 1 P 33/02	
A 6 1 P 35/00	A 6 1 P 33/14	
A 6 1 P 35/02	A 6 1 P 35/00	
A 6 1 P 37/00	A 6 1 P 35/02	
A 6 1 P 39/00	A 6 1 P 37/00	
A 6 1 P 41/00	A 6 1 P 39/00	
A 6 1 P 43/00	A 6 1 P 41/00	
C 0 7 K 14/705	A 6 1 P 43/00	1 1 1
C 0 7 K 16/28	C 0 7 K 14/705	
C 1 2 N 1/15	C 0 7 K 16/28	
C 1 2 N 1/19	C 1 2 N 1/15	
C 1 2 N 1/21	C 1 2 N 1/19	
C 1 2 N 5/10	C 1 2 N 1/21	
C 1 2 P 21/02	C 1 2 P 21/02	C
C 1 2 Q 1/68	C 1 2 Q 1/68	A
G 0 1 N 33/15	G 0 1 N 33/15	Z
G 0 1 N 33/50	G 0 1 N 33/50	Z
G 0 1 N 33/53	G 0 1 N 33/53	M
G 0 1 N 33/566	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 60/177,331

(32)優先日 平成12年1月21日(2000.1.21)

(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 ボーグン、マライア・アール

アメリカ合衆国カリフォルニア州9 4 5 7 7・サンレアンドロ・サンティアゴロード 1 4 2 4 4

(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・プリズベン・ゴールデンイーグルレーン 2 3 3

(72)発明者 ヤング、ジュンミン

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 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 レディ、ルーパ

アメリカ合衆国カリフォルニア州 9 4 0 8 6 ・ サニーベイル ・ # 3 ・ ウェストマッキンレーアベニ
ュー 1 2 3 3

F ターム(参考) 2G045 AA25 AA40 BA11 BA13 BB20 BB50 CB01 CB20 CB21 DA12
DA13 DA14 DA36 DA77 FB02 FB04 FB07
4B024 AA01 AA11 BA63 CA01 DA02 FA02 GA11 HA12
4B063 QA01 QA18 QQ02 QQ42 QR08 QR32 QR42 QR55 QS25 QS34
QX02
4B064 AG20 AG27 CA19 CC24 DA01 DA13
4B065 AA93Y AB01 AC14 BA02 CA24 CA25 CA44 CA46
4C084 AA02 AA07 AA13 AA17 BA01 BA08 BA22 BA23 CA18 DC50
MA01 NA14 ZA011 ZA021 ZA051 ZA061 ZA071 ZA081 ZA151 ZA161
ZA181 ZA201 ZA221 ZA241 ZA331 ZA361 ZA401 ZA421 ZA441 ZA451
ZA511 ZA551 ZA591 ZA661 ZA671 ZA681 ZA701 ZA711 ZA721 ZA731
ZA751 ZA811 ZA891 ZA941 ZA961 ZA971 ZB011 ZB051 ZB071 ZB111
ZB131 ZB151 ZB211 ZB261 ZB271 ZB331 ZB351 ZB371 ZB381 ZC011
ZC061 ZC311 ZC351 ZC371 ZC412 ZC551
4C086 AA01 AA03 EA16 MA01 MA02 MA03 MA04 MA05 MA10 NA14
ZA01 ZA02 ZA05 ZA06 ZA07 ZA08 ZA15 ZA16 ZA18 ZA20
ZA22 ZA24 ZA33 ZA36 ZA40 ZA42 ZA44 ZA45 ZA51 ZA55
ZA59 ZA66 ZA67 ZA68 ZA70 ZA71 ZA72 ZA73 ZA75 ZA81
ZA89 ZA94 ZA96 ZA97 ZB01 ZB05 ZB07 ZB11 ZB13 ZB15
ZB21 ZB26 ZB27 ZB33 ZB35 ZB37 ZB38 ZC01 ZC06 ZC31
ZC35 ZC37 ZC41 ZC55
4H045 AA10 AA11 BA10 CA40 DA50 EA20 EA50 FA74

专利名称(译)	G蛋白偶联受体		
公开(公告)号	JP2004500808A	公开(公告)日	2004-01-15
申请号	JP2001543585	申请日	2000-12-07
[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	バーフォードニール ボーゲンマライアアール オウヤングジャニス ヤングジュンミン リュデュングアイナエム レディルーバ		
发明人	バーフォード、ニール ボーゲン、マライア・アール オウ・ヤング、ジャニス ヤング、ジュンミン リュ、デュング・アイナ・エム レディ、ルーバ		
IPC分类号	A01K67/027 A61K31/711 A61K38/00 A61K45/00 A61K48/00 A61P1/00 A61P1/02 A61P1/04 A61P1/06 A61P1/08 A61P1/10 A61P1/14 A61P1/16 A61P1/18 A61P3/00 A61P3/04 A61P3/10 A61P5/18 A61P5 /38 A61P7/00 A61P7/06 A61P7/10 A61P9/00 A61P9/04 A61P9/10 A61P9/12 A61P9/14 A61P11/00 A61P11/06 A61P13/02 A61P13/08 A61P13/10 A61P13/12 A61P15/00 A61P15/08 A61P17/00 A61P17 /06 A61P17/16 A61P19/00 A61P19/02 A61P19/04 A61P19/10 A61P21/00 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 A61P31/12 A61P31/16 A61P31/18 A61P31/22 A61P33/00 A61P33/02 A61P33/14 A61P35/00 A61P35/02 A61P37/00 A61P39/00 A61P41/00 A61P43/00 C07K14 /47 C07K14/705 C07K14/78 C07K16/28 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N15/09 C12P21 /02 C12Q1/68 G01N33/15 G01N33/50 G01N33/53 G01N33/566		
CPC分类号	A01K2217/05 A61K38/00 A61P1/00 A61P1/02 A61P1/04 A61P1/06 A61P1/08 A61P1/10 A61P1/14 A61P1/16 A61P1/18 A61P3/00 A61P3/04 A61P11/00 A61P11/06 A61P13/02 A61P13/08 A61P13/10 A61P13/12 A61P15/00 A61P15/08 A61P17/00 A61P17/06 A61P17/16 A61P19/00 A61P19/02 A61P19 /04 A61P19/10 A61P21/00 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 /12 A61P31/16 A61P31/18 A61P31/22 A61P33/00 A61P33/02 A61P33/14 A61P35/00 A61P35/02 C07K14/47 C07K14/705 C07K14/78		
FI分类号	C12N15/00.ZNA.A A01K67/027 A61K31/711 A61K45/00 A61K48/00 A61P1/00 A61P1/02 A61P1/04 A61P1/06 A61P1/08 A61P1/10 A61P1/14 A61P1/16 A61P1/18 A61P3/00 A61P3/04 A61P3/10 A61P5 /18 A61P5/38 A61P7/00 A61P7/06 A61P7/10 A61P9/00 A61P9/04 A61P9/10 A61P9/12 A61P9/14 A61P11/00 A61P11/06 A61P13/02 A61P13/08 A61P13/10 A61P13/12 A61P15/00 A61P15/08 A61P17 /00 A61P17/06 A61P17/16 A61P19/00 A61P19/02 A61P19/04 A61P19/10 A61P21/00 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 A61P31/12 A61P31/16 A61P31/18 A61P31/22 A61P33/00 A61P33/02 A61P33/14 A61P35/00 A61P35/02 A61P37 /00 A61P39/00 A61P41/00 A61P43/00.111 C07K14/705 C07K16/28 C12N1/15 C12N1/19 C12N1/21 C12P21/02.C C12Q1/68.A G01N33/15.Z G01N33/50.Z G01N33/53.M G01N33/566 C12N5/00.A A61K37/02		

F-TERM分类号

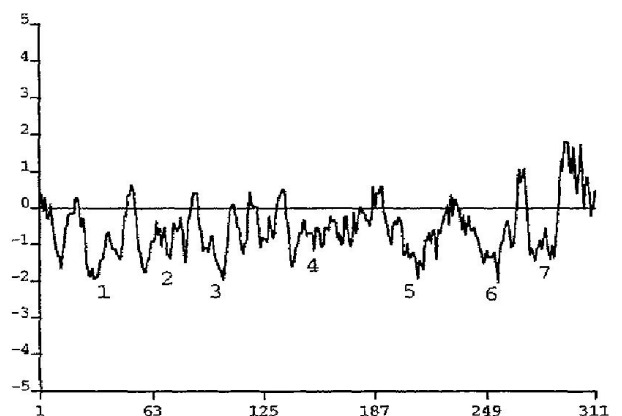
2G045/AA25 2G045/AA40 2G045/BA11 2G045/BA13 2G045/BB20 2G045/BB50 2G045/CB01 2G045/CB20 2G045/CB21 2G045/DA12 2G045/DA13 2G045/DA14 2G045/DA36 2G045/DA77 2G045/FB02 2G045/FB04 2G045/FB07 4B024/AA01 4B024/AA11 4B024/BA63 4B024/CA01 4B024/DA02 4B024/FA02 4B024/GA11 4B024/HA12 4B063/QA01 4B063/QA18 4B063/QQ02 4B063/QQ42 4B063/QR08 4B063/QR32 4B063/QR42 4B063/QR55 4B063/QS25 4B063/QS34 4B063/QX02 4B064/AG20 4B064/AG27 4B064/CA19 4B064/CC24 4B064/DA01 4B064/DA13 4B065/AA93Y 4B065/AB01 4B065/AC14 4B065/BA02 4B065/CA24 4B065/CA25 4B065/CA44 4B065/CA46 4C084/AA02 4C084/AA07 4C084/AA13 4C084/AA17 4C084/BA01 4C084/BA08 4C084/BA22 4C084/BA23 4C084/CA18 4C084/DC50 4C084/MA01 4C084/NA14 4C084/ZA011 4C084/ZA021 4C084/ZA051 4C084/ZA061 4C084/ZA071 4C084/ZA081 4C084/ZA151 4C084/ZA161 4C084/ZA181 4C084/ZA201 4C084/ZA221 4C084/ZA241 4C084/ZA331 4C084/ZA361 4C084/ZA401 4C084/ZA421 4C084/ZA441 4C084/ZA451 4C084/ZA511 4C084/ZA551 4C084/ZA591 4C084/ZA661 4C084/ZA671 4C084/ZA681 4C084/ZA701 4C084/ZA711 4C084/ZA721 4C084/ZA731 4C084/ZA751 4C084/ZA811 4C084/ZA891 4C084/ZA941 4C084/ZA961 4C084/ZA971 4C084/ZB011 4C084/ZB051 4C084/ZB071 4C084/ZB111 4C084/ZB131 4C084/ZB151 4C084/ZB211 4C084/ZB261 4C084/ZB271 4C084/ZB331 4C084/ZB351 4C084/ZB371 4C084/ZB381 4C084/ZC011 4C084/ZC061 4C084/ZC311 4C084/ZC351 4C084/ZC371 4C084/ZC412 4C084/ZC551 4C086/AA01 4C086/AA03 4C086/EA16 4C086/MA01 4C086/MA02 4C086/MA03 4C086/MA04 4C086/MA05 4C086/MA10 4C086/NA14 4C086/ZA01 4C086/ZA02 4C086/ZA05 4C086/ZA06 4C086/ZA07 4C086/ZA08 4C086/ZA15 4C086/ZA16 4C086/ZA18 4C086/ZA20 4C086/ZA22 4C086/ZA24 4C086/ZA33 4C086/ZA36 4C086/ZA40 4C086/ZA42 4C086/ZA44 4C086/ZA45 4C086/ZA51 4C086/ZA55 4C086/ZA59 4C086/ZA66 4C086/ZA67 4C086/ZA68 4C086/ZA70 4C086/ZA71 4C086/ZA72 4C086/ZA73 4C086/ZA75 4C086/ZA81 4C086/ZA89 4C086/ZA94 4C086/ZA96 4C086/ZA97 4C086/ZB01 4C086/ZB05 4C086/ZB07 4C086/ZB11 4C086/ZB13 4C086/ZB15 4C086/ZB21 4C086/ZB26 4C086/ZB27 4C086/ZB33 4C086/ZB35 4C086/ZB37 4C086/ZB38 4C086/ZC01 4C086/ZC06 4C086/ZC31 4C086/ZC35 4C086/ZC37 4C086/ZC41 4C086/ZC55 4H045/AA10 4H045/AA11 4H045/BA10 4H045/CA40 4H045/DA50 4H045/EA20 4H045/EA50 4H045/FA74

优先权
 60/172852 1999-12-10 US
 60/171732 1999-12-22 US
 60/176148 2000-01-14 US
 60/177331 2000-01-21 US

外部链接 [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。



SEQ ID NO:1
 (Incyte ID No. 104941CD1)