

(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2004-523203

(P2004-523203A)

(43) 公表日 平成16年8月5日(2004.8.5)

(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
A 6 1 K 31/7088	A 6 1 K 31/7088	2 G O 4 5
A 6 1 K 38/00	A 6 1 K 39/395	4 B O 2 4
A 6 1 K 39/395	A 6 1 K 39/395	D
A 6 1 K 45/00	A 6 1 K 45/00	N
		4 B O 6 3
		4 B O 6 4
		4 B O 6 5
	審査請求 未請求 予備審査請求 有	(全 253 頁) 最終頁に続く

(21) 出願番号	特願2002-504306 (P2002-504306)	(71) 出願人	301005050
(86) (22) 出願日	平成13年6月15日 (2001.6.15)		インサイト・ゲノミックス・インコーポレイテッド
(85) 翻訳文提出日	平成14年12月13日 (2002.12.13)		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) 国際出願番号	PCT/US2001/019275	(74) 代理人	100089266
(87) 国際公開番号	W02001/098351		弁理士 大島 陽一
(87) 国際公開日	平成13年12月27日 (2001.12.27)	(72) 発明者	ラル、ブリーティ
(31) 優先権主張番号	60/212, 483		アメリカ合衆国カリフォルニア州95056・サンタクララ・ビーオーボックス 5142
(32) 優先日	平成12年6月16日 (2000.6.16)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/213, 954		
(32) 優先日	平成12年6月22日 (2000.6.22)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/215, 209		
(32) 優先日	平成12年6月29日 (2000.6.29)		
(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 (G C R E C) およびG C R E Cを同定し、コードするポリヌクレオチドを提供する。本発明はまた、発現ベクター、宿主細胞、抗体、アゴニストおよびアンタゴニストをも提供する。本発明はまた、G C R E Cの異常発現に関連する疾患を診断、治療または予防する方法をも提供する。

【特許請求の範囲】

【請求項 1】

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

(a) SEQ ID NO: 1 - 17 を有する群から選択したアミノ酸配列を含むポリペプチド

(b) SEQ ID NO: 1 - 17 からなる群から選択したアミノ酸配列と少なくとも 90% が同一であるような天然のアミノ酸配列を含むポリペプチド

(c) SEQ ID NO: 1 - 17 を有する群から選択したアミノ酸配列を有するポリペプチドの生物学的活性断片

(d) SEQ ID NO: 1 - 17 を有する群から選択したアミノ酸配列を有するポリペプチドの免疫抗原性断片 10

【請求項 2】

SEQ ID NO: 1 - 17 を有する群から選択した請求項 1 に記載の単離されたポリペプチド。

【請求項 3】

請求項 1 のポリペプチドをコードする単離されたポリヌクレオチド。

【請求項 4】

請求項 2 のポリペプチドをコードする単離されたポリヌクレオチド。

【請求項 5】

SEQ ID NO: 18 - 34 を有する群から選択した請求項 4 に記載の単離されたポリヌクレオチド。 20

【請求項 6】

請求項 3 に記載のポリヌクレオチドに機能的に結合したプロモーター配列を含む組換えポリヌクレオチド。

【請求項 7】

請求項 6 に記載の組換えポリヌクレオチドを用いて形質転換した細胞。

【請求項 8】

請求項 6 に記載の組換えポリヌクレオチドを含む遺伝形質転換体。

【請求項 9】

請求項 1 のポリペプチドを生産する方法であって、以下の過程を含む方法。 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過程と、

(b) そのように発現した前記ポリペプチドを回収する過程。

【請求項 10】

請求項 1 に記載のポリペプチドと特異的に結合する単離された抗体。

【請求項 11】

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

(a) SEQ ID NO: 18 - 34 を有する群から選択したポリヌクレオチド配列を有するポリヌクレオチド 40

(b) SEQ ID NO: 18 - 34 からなる群から選択したポリヌクレオチド配列と少なくとも 90% が同一であるような天然のポリヌクレオチドを含むポリヌクレオチド

(c) (a) のポリヌクレオチドに相補的なポリヌクレオチド

(d) (b) のポリヌクレオチドに相補的なポリヌクレオチド

(e) (a) ~ (d) の RNA 等価物

【請求項 12】

請求項 11 に記載のポリヌクレオチドの少なくとも 60 の連続したヌクレオチドを含む単離されたポリヌクレオチド。

【請求項 13】

請求項 11 に記載のポリヌクレオチドの配列を有する標的ポリヌクレオチドをサンプル中 50

から検出する方法であって、

(a) 前記サンプル中の前記標的ポリヌクレオチドに相補的な配列を含む少なくとも 2 0 の連続したヌクレオチドを含むプローブを用いて前記サンプルをハイブリダイズする過程 (ただし、前記プローブと前記標的ポリヌクレオチドまたは断片の間でハイブリダイゼーション複合体が形成されるような条件下で、前記プローブが前記標的ポリヌクレオチドに特異的にハイブリダイズする) と、

(b) 前記ハイブリダイゼーション複合体の存在・不存在を検出し、該複合体が存在する場合にはオプションでその量を検出する過程を含むことを特徴とする方法。

【請求項 1 4】

前記プローブが少なくとも 6 0 の連続したヌクレオチドを含むことを特徴とする請求項 1 3 に記載の方法。 10

【請求項 1 5】

請求項 1 1 に記載のポリヌクレオチドの配列を有する標的ポリヌクレオチドをサンプル中から検出する方法であって、

(a) ポリメラーゼ連鎖反応増幅を用いて前記標的ポリヌクレオチドまたはその断片を増幅する過程と、

(b) 前記標的ポリヌクレオチドまたはその断片の存在・不存在を検出し、該標的ポリヌクレオチドまたはその断片が存在する場合にはオプションでその量を検出する過程を含むことを特徴とする方法。

【請求項 1 6】

請求項 1 のポリペプチドと、薬剤として許容できる賦形剤とを含む組成物。 20

【請求項 1 7】

前記ポリペプチドが、SEQ ID NO : 1 - 1 7 を有する群から選択したアミノ酸配列を含むことを特徴とする請求項 1 6 に記載の成分。

【請求項 1 8】

機能性 G C R E C の発現低下に関連する疾患又は病状を治療する方法であって、そのような治療を必要とする患者に対して請求項 1 6 に記載の成分を投与する過程を含むことを特徴とする方法。

【請求項 1 9】

請求項 1 に記載のポリペプチドのアゴニストとして有効性を確認するために化合物をスクリーニングする方法であって、 30

(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過程と、

(b) 前記サンプルにおいてアゴニスト活性を検出する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 に記載の方法によって同定したアゴニスト化合物と薬剤として許容できる賦形剤とを含む組成物。

【請求項 2 1】

機能性 G C R E C の発現低下に関連する疾患又は病状を治療する方法であって、そのような治療を必要とする患者に対して請求項 2 0 に記載の成分を投与する過程を含むことを特徴とする方法。 40

【請求項 2 2】

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

(a) 請求項 1 のポリペプチドを含むサンプルを化合物に曝露する過程と、

(b) 前記サンプルにおいてアンタゴニスト活性を検出する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 3】

請求項 2 2 に記載の方法によって同定したアンタゴニスト化合物と、薬剤として許容できる賦形剤とを含む組成物。

【請求項 24】

機能性 G C R E C の過剰発現に関連する疾患又は病状の治療方法であって、そのような治療を必要とする患者に対して請求項 23 に記載の成分を投与する過程を含むことを特徴とする方法。

【請求項 25】

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

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

(b) 請求項 1 に記載のポリペプチドの試験化合物との結合を検出し、それによって請求項 1 に記載のポリペプチドに特異結合する化合物を同定する過程とを含むことを特徴とする方法。 10

【請求項 26】

請求項 1 に記載のポリペプチドの活性を調節する化合物をスクリーニングする方法であって、

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

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

(c) 試験化合物の存在下での請求項 1 に記載のポリペプチドの活性を、試験化合物の不存在下での請求項 1 に記載のポリペプチドの活性と比較する過程とを含み、試験化合物の存在下での請求項 1 に記載のポリペプチドの活性の変化が、請求項 1 に記載のポリペプチドの活性を調節する化合物を標示することを特徴とする方法。 20

【請求項 27】

請求項 5 の配列を含む標的ポリヌクレオチドの発現を変化させるのに効果的な化合物をスクリーニングする方法であって、

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

(b) 前記標的ポリヌクレオチドの変異発現を検出する過程と、

(c) 可変量の前記化合物の存在下と前記化合物の不存在下で、前記標的ポリヌクレオチドの発現を比較する過程とを含むことを特徴とする方法。 30

【請求項 28】

試験化合物の毒性を算定する方法であって、

(a) 核酸を含む生物学的サンプルを前記試験化合物で処理する過程と、

(b) 処理した前記生体サンプルの核酸と、請求項 11 のポリヌクレオチドの少なくとも 20 の連続するヌクレオチドを含むプローブをハイブリダイズさせるステップであって、このハイブリダイゼーションゼーションが、前記プローブと前記生体サンプルの標的ポリヌクレオチドとの間で特異的なハイブリダイゼーション複合体が形成される条件下で行われ、前記標的ポリヌクレオチドが、請求項 11 のポリヌクレオチドのポリヌクレオチド配列またはその断片を含むポリヌクレオチドである、前記ステップと、

(c) ハイブリダイゼーション複合体の収量を定量するステップと、 40

(d) 前記処理された生物学的サンプル中の前記ハイブリタイゼーション複合体の量を、処理されていない生物学的サンプル中の前記ハイブリタイゼーション複合体の量と比較する過程とを含み、前記処理された生物学的サンプル中の前記ハイブリタイゼーション複合体の量の差が、前記試験化合物の毒性を標示することを特徴とする方法。

【請求項 29】

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

(a) 前記抗体が前記ポリペプチドに結合し、抗体とポリペプチドとの複合体が形成されるのに適した条件下で、前記生物学的サンプルを請求項 10 に記載の抗体と結合する過程と、

(b) 前記複合体を検出する過程とを含み、前記複合体の存在が、前記生物学的サンプル中の前記ポリペプチドの存在と相関することを特徴とする方法。

【請求項30】

前記抗体が、

(a) キメラ抗体

(b) 単鎖抗体

(c) Fab断片

(d) F(ab')₂断片

(e) ヒト化抗体のいずれかであることを特徴とする請求項10に記載の抗体。

【請求項31】

請求項10に記載の抗体と、許容できる賦形剤とを含む化合物。

【請求項32】

被検者のGCRECの発現に関連する病状又は疾患の診断方法であって、請求項31に記載の化合物の有効量を前記被検者に投与する過程を含むことを特徴とする方法。

【請求項33】

前記抗体が標識されることを特徴とする請求項31に記載の化合物。

【請求項34】

被検者のGCRECの発現に関連する病状又は疾患の診断方法であって、請求項33に記載の化合物の有効量を前記被検者に投与する過程を含むことを特徴とする方法。

【請求項35】

請求項10に記載の抗体の特異性を有するポリクローナル抗体を調製する方法であって、(a) 抗体反応を誘発する条件下で、SEQ ID NO: 1-17を有する群から選択したアミノ酸配列含むポリペプチドまたはその免疫抗原性断片を用いて動物を免疫化する過程と、

(b) 前記動物から抗体を単離する過程と、

(c) 前記単離された抗体をポリペプチドでスクリーニングし、それによって、SEQ ID NO: 1-17を有する群から選択したアミノ酸配列を有するポリペプチドに特異結合するようなポリクローナル抗体を同定する過程とを含むことを特徴とする方法。

【請求項36】

請求項35に記載の方法で産出した抗体。

【請求項37】

請求項36に記載の抗体及び適切なキャリアを含む化合物。

【請求項38】

請求項10に記載の抗体の特異性を有するモノクローナル抗体を製造する方法であって、(a) 抗体反応を誘発する条件下で、SEQ ID NO: 1-17を有する群から選択したアミノ酸配列含むポリペプチドまたはその免疫抗原性断片を用いて動物を免疫化する過程と、

(b) 前記動物から抗体産出細胞を単離する過程と、

(c) 前記抗体産出細胞を不死化の細胞と融合して、モノクローナル抗体を産出するハイブリドーマ細胞を形成する過程と、

(d) 前記ハイブリドーマ細胞を培養する過程と、

(e) SEQ ID NO: 1-17を有する群から選択したアミノ酸配列を有するポリペプチドに特異結合するような前記培養モノクローナル抗体から単離する過程とを含むことを特徴とする方法。

【請求項39】

請求項38に記載の方法で産出したモノクローナル抗体。

【請求項40】

請求項39に記載の抗体及び適切なキャリアを含む化合物。

【請求項41】

Fab発現ライブラリのスクリーニングにより前記抗体を産出することを特徴とする請求

10

20

30

40

50

項 10 に記載の抗体。

【請求項 42】

組換え免疫グロブリンライブラリのスクリーニングにより前記抗体を産出することを特徴とする請求項 10 に記載の抗体。

【請求項 43】

SEQ ID NO: 1 - 17 を有する群から選択したアミノ酸配列を有するポリペプチドを検出する方法であって、

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

(b) 特異結合を検出する過程とを含み、該特異結合が、SEQ ID NO: 1 - 17 を有する群から選択したアミノ酸配列を有するポリペプチドがサンプル中に存在することを標示することを特徴とする方法。 10

【請求項 44】

SEQ ID NO: 1 - 17 を有する群から選択したアミノ酸配列を有するポリペプチドを精製する方法であって、

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

(b) 前記サンプルから前記抗体を分離し、SEQ ID NO: 1 - 17 を有する群から選択したアミノ酸配列を有する精製ポリペプチドを得る過程とを含むことを特徴とする方法。 20

【請求項 45】

SEQ ID NO: 1 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 46】

SEQ ID NO: 2 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 47】

SEQ ID NO: 3 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 48】

SEQ ID NO: 4 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 49】

SEQ ID NO: 5 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。 30

【請求項 50】

SEQ ID NO: 6 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 51】

SEQ ID NO: 7 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 52】

SEQ ID NO: 8 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 53】

SEQ ID NO: 9 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 54】

SEQ ID NO: 10 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。 40

【請求項 55】

SEQ ID NO: 11 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 56】

SEQ ID NO: 12 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 57】

SEQ ID NO: 13 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 58】

SEQ ID NO: 14 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 59】

SEQ ID NO: 15 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。 50

【請求項 60】

SEQ ID NO : 16 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 61】

SEQ ID NO : 17 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 62】

SEQ ID NO : 18 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

【請求項 63】

SEQ ID NO : 19 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

【請求項 64】

SEQ ID NO : 20 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

【請求項 65】

SEQ ID NO : 21 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

【請求項 66】

SEQ ID NO : 22 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

【請求項 67】

SEQ ID NO : 23 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

【請求項 68】

SEQ ID NO : 24 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

【請求項 69】

SEQ ID NO : 25 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

【請求項 70】

SEQ ID NO : 26 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

【請求項 71】

SEQ ID NO : 27 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

【請求項 72】

SEQ ID NO : 28 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

【請求項 73】

SEQ ID NO : 29 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

【請求項 74】

SEQ ID NO : 30 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

【請求項 75】

SEQ ID NO : 31 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

【請求項 76】

SEQ ID NO : 32 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

【請求項 77】

SEQ ID NO : 33 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

10

20

30

40

50

SEQ ID NO : 33 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

【請求項 78】

SEQ ID NO : 34 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

【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し、細胞増殖異常、神経障害、心血管障害、胃腸障害、自己免疫/炎症疾患及び代謝障害並びにウイルス感染症の診断、治療並びに予防におけるこれらの配列の利用と、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因性化合物の効果の算定におけるこれらの配列の利用に関する。

10

【0002】

(発明の背景)

シグナル伝達は、それによって細胞が細胞外シグナルに応答するような一般的プロセスである。原形質膜でのシグナル伝達は、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から開始する。そのようにして活性化された受容体は、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カスケードを誘発する。シグナル伝達のこのプロセスは、細胞増殖、分化及び遺伝子転移を含む全てのタイプの細胞機能を制御する。同定された遺伝子の最大ファミリーの1つによりコードされたGタンパク質結合受容体(GPCR)は、原形質膜での細胞外シグナルの伝達において中心的役割を果たす。GPCRが治療の標的として成功であることは歴史的に証明されている。

20

【0003】

GPCRは、7つの疎水性膜貫通ドメインの存在により特徴付けられた膜内在性タンパク質であり、それらのドメインがまとまって逆平行アルファ()螺旋の束を形成するようになる。GPCRのサイズは、400以下から1000以上のアミノ酸に及ぶ(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、細胞外にあり、長さが可変で、多くの場合グリコシル化される。カルボキシル末端は、細胞質内にあり、通常はリン酸化される。細胞外ループは、細胞内ループと交互に現れ、膜貫通ドメインを結合する。第2及び第3細胞外ループを結合するシステインジスルフィド架橋は、アゴニスト及びアンタゴニストと相互に作用し得る。GPCRの最も保存されたドメインは、膜貫通ドメイン及び最初の2つの細胞質ループである。膜貫通ドメインは、受容体の構造的及び機能的特徴をある程度明らかにする。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末端セグメント即ち1個若しくは数個の第3細胞外ループは、リガンド結合にも関与し得る。リガンド結合は、受容体の細胞内部分で構造変化を誘導することにより受容体を活性化する。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ヘテロ三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. 及び S. Arkininstall (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego CA, 2-6 ページ、Bolander, F. F. (1994) Molecular Endocrinology, Academic Press, San Diego CA, 162-176 ページ; Baldwin, J. M. (1994) Curr. Opin. Cell Biol. 50

30

40

50

6 : 1 8 0 1 9 0 等参照)。

【 0 0 0 4 】

GPCRには、感覚性シグナルメディエータ(例えば光及び嗅覚刺激性分子)の受容体、アデノシン、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及びバソプレシン、生体アミン(例えばドーパミン、エピネフリン及びノルエピネフリン、ヒスタミン、グルタミン酸(代謝調節作用)、アセチルコリン(ムスカリン様作用)及びセロトニン)、ケモカイン、炎症の脂質メディエータ(例えばプロスタグランジン及びプロスタノイド、血小板活性化因子及びロイコトリエン)及びペプチドホルモン(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)、ゴナドトロピン放出ホルモン(GnRH)、ニューロキニン及び甲状腺刺激ホルモン放出ホルモン(TRH)及びオキシトシン)がある。同定された刺激に対する受容体として作用するGPCRは、オルファン受容体として知られている。

10

【 0 0 0 5 】

GPCRファミリーの多様性は、別のスプライシングによって更に増加する。多くのGPCR遺伝子にはイントロンが含まれ、スプライシング変異体が同定されるそのような受容体は現在では30を超えている。変異の数が最も大きいのは、タンパク質C末端においてである。細胞外ループまたは膜貫通ドメインは頻度が低いのに対して、N末端及び細胞質内ループ変異体もまた高頻度である。変異が発生し得るような部位を2ヶ所以上有する受容体もある。スプライシング変異体は、分布、シグナル伝達、結合、制御及びリガンド結合の特徴に観察される差異に基づき、機能的に異なるようである(Kilpatrick, G. J. ら(1999) Trends Pharmacol. Sci. 20: 294-301)。

20

【 0 0 0 6 】

GPCRは、3つの主要なサブファミリー即ちロドプシン様、セクレチン様及び代謝調節型グルタミン酸受容体サブファミリーに分けることができる。GPCRサブファミリーのメンバーは、同様の機能及び特徴的な膜7回貫通構造を共有しているが、アミノ酸配列は互いに異なる。最大のファミリーはロドプシン様GPCRを有し、ロドプシン様GPCRはホルモン、神経伝達物質及び光を含む多様な細胞外シグナルを伝送する。ロドプシンは、動物の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体(杆体)細胞に見られる膜性スタックに包埋されている。各ロドプシン分子は、原形質膜ナトリウムチャンネルの閉鎖に導くcGMPレベルの低下を誘発することにより光の光子に反応する。この方法で、可視シグナルが神経インパルスに変換される。その他のロドプシン様GPCRは、神経伝達物質への反応に直接関与する。このようなGPCRには、アドレナリンに対する受容体(アドレナリン受容体)、アセチルコリンに対する受容体(ムスカリン様受容体)、アデノシンに対する受容体、ガラニンに対する受容体及びグルタミン酸に対する受容体(N-メチル-D-アスパラギン酸/NMDA受容体)がある(Watson, S. 及びS. Arkininstall(1994) The G-Protein Linked Receptor Facts Book, Academic Press, San Diego CA, 7-9, 19-22, 32-35, 130-131, 214-216, 221-222の各ページ、Habert-Ortoli, E. ら(1994) Proc. Natl. Acad. Sci. USA 91: 9780-9783に概説されている)。

30

40

【 0 0 0 7 】

ガラニン受容体は、インスリン、アセチルコリン、セロトニン及びノルアドレナリンの分泌を阻害する神経内分泌ペプチドガラニンの活性を仲介し、プロラクチン及び成長ホルモンの放出を刺激する。ガラニン受容体は、摂食に関する疾患、痛み、うつ病及びアルツハイマー病のフィードに関与している(Kask, K. ら(1997) Life Sci. 60: 1523-1533)。その他の神経系ロドプシン様GPCRは、発達及び

50

神経病理学において役割を有すると考えられる、リゾホスファチジン酸その他のリゾホスファチドに対する受容体など、増大する数のファミリーを含む (Chun, J. ら (1999) *Cell Biochem. Biophys.* 30: 213-242)。

【0008】

GPCRの最大のサブファミリーである嗅覚受容体もまた、ロドプシン様GPCRファミリーのメンバーである。この受容体は、臭覚シグナルの伝達により機能する。異なる臭気を区別するためには、複数の異なる嗅覚受容体が必要である。各嗅覚感覚ニューロンは1種類の嗅覚受容体しか発現せず、特有の受容体を発現するニューロンが、鼻の経路中の異なる位置を占める。例えば、ラット脳ライブラリから単離したRA1c受容体は、脳の非常に特異な領域及び画定された嗅覚上皮の特定な区域だけに発現が限定されていることを示していた (Raining, 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ない。例えば、典型的なGPCR特性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、味覚及び嗅覚組織のみならず男性生殖組織においても発現パターンを示す (Thomas, M. B. ら (1996) *Gene* 178: 1-5)。

10

【0009】

セクレチン様GPCRサブファミリーのメンバーは、セクレチン、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、及び血管作用性小腸ペプチドなどのペプチドホルモンをリガンドとして有する。例えば、セクレチン受容体は、膵臓及び小腸で酵素及びイオンの分泌を刺激するペプチドホルモンであるセクレチンに対応する (前出のWatson, 278-283ページ)。セクレチン受容体は、長さが約450アミノ酸であり、胃腸細胞の原形質膜に見られる。セクレチンがその受容体に結合することにより、cAMPの産出が誘発される。

20

【0010】

炎症及び免疫反応に結びつけられるセクレチン様GPCRの例には、EGFのモジュールを含んだムチン様ホルモン受容体 (Emr1) 及び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。これら膜7回貫通ホルモン受容体は、*in vivo*でヘテロ二量体として存在し、3から7の潜在的カルシウム結合EGF様モチーフを含む。CD97は、主に白血球で発現し、活性化されたB及びT細胞上で顕著に上方制御される (McKnight, A. J. 及びS. Gordon (1998) *J. Leukoc. Biol.* 63: 271-280)。

30

【0011】

第3GPCR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、中枢神経系の主要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、細胞内エフェクターの活性を調節し、長期の増強作用に関与する (前出のWatson, p. 130)。Ca²⁺ 検出受容体は、カルシウムイオンの細胞外濃度の変化を検出するものであり、カルシウム結合に関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には、フェロモン受容体、GABA_B受容体及び味覚受容体もある。

40

【0012】

その他のGPCRのサブファミリーには、線虫及びCaenorhabditis briggsaeに見られる化学受容体遺伝子の2つの群があり、これらは哺乳動物の嗅覚受容体遺伝子に直接関係している。細胞膜上の接合因子への反応に関与している酵母フェロモン受容体STE2及びSTE3は、細胞性粘菌で、個々の細胞の集合を調整し、複数の発達の制御遺伝子の発現を制御すると考えられているcAMP受容体と同様に、膜7回貫通シグネチャを有する。

【0013】

GPCR突然変異は、機能または恒常的活性化の減少を招き得るものであり、多数のヒト疾患に関係してきた (前出のCoughlin)。例えば、色素性網膜炎はロドプシン遺

50

伝子の突然変異から発生し得る。更に、甲状腺刺激ホルモン受容体の体細胞活性化突然変異は、機能亢進甲状腺アデノーマの原因となることが報告されており、恒常的活性化に感受性が高い或るGPCRが癌原遺伝子のように振舞い得ることを示唆している (Parma, J.ら (1993) Nature 365: 649-651)。以下のリガンド即ち黄体形成ホルモン (思春期早発症)、バソプレシンV₂ (X連鎖の腎原発性糖尿病)、グルカゴン (糖尿病及び高血圧)、カルシウム (副甲状腺機能亢進症、低カルシウム尿症 (hypocalcuria)、高カルシウム血症)、副甲状腺ホルモン (短肢小人症)、 β_3 -アドレナリン受容体 (肥満症、非インスリン依存型糖尿病)、成長ホルモン放出ホルモン (小人症) 及び副腎皮質刺激ホルモン (糖質コルチコイド欠乏症) に対するGPCR受容体には、ヒトの疾患に関連する突然変異も含まれる (Wilson, S.ら (1998) Br. J. Pharmacol. 125: 1387-1392、Stadel, J.M.ら (1997) Trends Pharmacol. Sci. 18: 430-437)。GPCRは、抑うつ症、分裂病、不眠症、高血圧、不安、ストレス、腎不全その他幾つかの心血管障害にも関与している (Horn, F.及びG. Vriend (1998) J. Mol. Med. 76: 464-468)。

【0014】

更に、GPCRの活性化及び阻害に直接作用する数百の新薬が過去20年のうちに認知されてきた。これらの薬の治療標的は、癌、骨粗鬆症、子宮内膜症のみならず心血管障害、胃腸障害、中枢神経系疾患を含めた幅広い疾病及び疾患に及ぶ (前出のWilson、同Stadel)。例えば、ドーパミンアゴニストのLドーパはパーキンソン病の治療に用いられ、ドーパミンアンタゴニストは精神分裂病及び初期段階のハンチントン病の治療に用いられる。アドレナリン受容体のアゴニスト及びアンタゴニストは、喘息、高血圧その他の心血管障害及び不安の治療に用いられてきた。ムスカリン様アゴニストは、緑内障及び頻脈の治療に用いられ、セロトニン5HT_{1D}アンタゴニストは片頭痛に対して用いられ、ヒスタミンH₁アンタゴニストはアレルギー性及びアナフィラキシー性反応、枯草熱、痒み及び動揺病に対して用いられる (前出のHorn)。

【0015】

最近の調査では、糖尿病、肥満症及び骨粗鬆症を含む代謝障害の治療においてGPCRを将来的に使用する可能性を示唆している。例えば、腎原発性糖尿病の原因となる突然変異体V₂バソプレシン受容体は、突然変異を含む領域に及ぶC末端V₂受容体ペプチドの同時発現により *in vitro* で機能的に救出され得る。これは疾病治療の新規なストラテジーを示唆する (Schoneberg, T.ら (1996) EMBO 1. 15: 1283-1291)。メラノコルチン4受容体 (MC4R) における突然変異は、ヒトの体重調節及び肥満症に結びつけられる。バソプレシンV₂受容体の突然変異体により、これらMC4Rの突然変異体は原形質膜への輸送に欠陥があるので (Ho, G及びR. G. MacKenzie (1999) J. Biol. Chem. 274: 35816-35822)、従って同様のストラテジーで処理し得る。副甲状腺ホルモン (PTH) のための1型受容体は、血流中のカルシウムホメオスタシスのPTH依存型制御を仲介するGPCRである。PTH/受容体の相互作用の研究は、骨粗鬆症の治療のための新規なPTH受容体リガンドの開発を可能にし得る (Mannstadt, M.ら (1999) Am. J. Physiol. 277: F665-F675)。

【0016】

GPCRのケモカイン受容体グループは、炎症及び感染症の治療に役立つ可能性を有する (Locati, M.及びP.M. Murphy (1999) Annu. Rev. Med. 50: 425-440の概説を参照)。ケモカインは、白血球輸送、造血及び血管形成の制御において細胞内シグナルとして作用する小ポリペプチドである。マウスにおける種々のケモカイン受容体の標的破壊によって、これらの受容体が異常炎症において及び多発性硬化症などの自己免疫異常において役割を果たしていることを示す。ヘルペスウイルス及びヒト免疫不全症ウイルス (HIV-1) を含む感染体が、感染を促進させるために、ケモカイン受容体を利用している。ケモカイン受容体CCR5は、HIV-1に

よるT細胞の感染に対する補助受容体として作用するものであり、その受容体で一部が切断されたものはAIDSに対する抵抗力を生じさせた。この結果は、CCR5のアンタゴニストがAIDSの予防に有用たり得ることを示唆する。

【0017】

新たなGタンパク質結合受容体及びそれをコードするポリヌクレオチドの発見は、細胞増殖異常、神経障害、心血管障害、胃腸障害、自己免疫/炎症疾患及び代謝障害並びにウイルス感染症の診断、治療並びに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影響についての評価にも有用である。

【0018】

10

(発明の概要)

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」、および「GCREC-17」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。或る実施態様において本発明は、(a)SEQ ID NO:1-17を有する群から選択したアミノ酸配列からなるポリペプチド、(b)SEQ ID NO:1-17を有する群から選択したアミノ酸配列と少なくとも90%が同一である天然のアミノ酸配列を有するポリペプチド、(c)SEQ ID NO:1-17を有する群から選択したアミノ酸配列を有するポリペプチドの生物学的活性断片、または(d)SEQ ID NO:1-17を有する群から選択したアミノ酸配列を有するポリペプチドの免疫抗原性断片を含む群から選択した実質上単離されたポリペプチドを提供する。一実施態様では、SEQ ID NO:1-17のアミノ酸配列を含む実質上単離されたポリペプチドを提供する。

20

【0019】

また、本発明は(a)SEQ ID NO:1-17を有する群から選択したアミノ酸配列からなるポリペプチド、(b)SEQ ID NO:1-17を有する群から選択したアミノ酸配列と少なくとも90%が同一である天然のアミノ酸配列からなるポリペプチド、(c)SEQ ID NO:1-17を有する群から選択したアミノ酸配列を有するポリペプチドの生物学的活性断片、または(d)SEQ ID NO:1-17を有する群から選択したアミノ酸配列を有するポリペプチドの免疫抗原性断片、を含む群から選択されたポリペプチドをコードする実質上単離されたポリヌクレオチドを提供する。一実施態様では、ポリヌクレオチドはSEQ ID NO:1-17を有する群から選択したポリペプチドをコードする。別の実施態様では、ポリヌクレオチドはSEQ ID NO:18-34を有する群から選択される。

30

【0020】

本発明は更に、(a)SEQ ID NO:1-17を有する群から選択したアミノ酸配列からなるポリペプチド、(b)SEQ ID NO:1-17を有する群から選択したアミノ酸配列と少なくとも90%の相同性のある天然のアミノ酸配列からなるポリペプチド、(c)SEQ ID NO:1-17を有する群から選択したアミノ酸配列を有するポリペプチドの生物学的活性断片、または(d)SEQ ID NO:1-17を有する群から選択したアミノ酸配列を有するポリペプチドの免疫抗原性断片を有する群から選択したポリペプチドをコードするようなポリヌクレオチドと機能的に結合したプロモーター配列を有する組換えポリヌクレオチドを提供する。一実施態様では、本発明は組換えポリヌクレオチドを用いて形質転換した細胞を提供する。別の実施態様では、本発明は組換えポリヌクレオチドを含む遺伝形質転換体を提供する。

40

【0021】

更に本発明は、(a)SEQ ID NO:1-17からなる一群から選択されたアミノ酸配列からなるポリペプチドと、(b)SEQ ID NO:1-17からなる一群から選択

50

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からなるポリペプチドと、(c)SEQ ID NO: 1-17からなる一群から選択されたアミノ酸配列を有するポリペプチドの生物学的に活性な断片と、(d)SEQ ID NO: 1-17とからなる一群から選択されたアミノ酸配列を有するポリペプチドの免疫原性断片とで構成される一群から選択されたポリペプチドを製造する方法を提供する。製造方法は、(a)組換えポリヌクレオチドを用いて形質転換した細胞をポリペプチドの発現に適した条件下で培養する過程と、(b)そのように発現したポリペプチドを受容する過程とを有し、組換えポリヌクレオチドはポリペプチドをコードするポリヌクレオチドに機能的に結合したプロモーター配列を有する。

【0022】

本発明は更に、(a)SEQ ID NO: 1-17を有する群から選択したアミノ酸配列を有するポリペプチド、(b)SEQ ID NO: 1-17を有する群から選択したアミノ酸配列と少なくとも90%同一である天然のアミノ酸配列を有するポリペプチド、(c)SEQ ID NO: 1-17を有する群から選択したアミノ酸配列を有するポリペプチドの生物学的活性断片、または(d)SEQ ID NO: 1-17を有する群から選択したアミノ酸配列を有するポリペプチドの免疫抗原性断片から構成される群から選択されたポリペプチドに特異結合するような実質上単離された抗体を提供する。

【0023】

本発明は更に、(a)SEQ ID NO: 18-34を有する群から選択したポリヌクレオチド配列を有するポリヌクレオチドと、(b)SEQ ID NO: 18-34を有する群から選択したポリヌクレオチド配列と少なくとも90%が同一である天然のポリヌクレオチド配列を有するポリヌクレオチドと、(c)(a)のポリヌクレオチドに相補的なポリヌクレオチドと、(d)(b)のポリヌクレオチドに相補的なポリヌクレオチドと、(e)(a)~(d)のRNA等価物とで構成される群から選択した実質上単離されたポリヌクレオチドを提供する。一実施態様では、ポリヌクレオチドは少なくとも60の連続したヌクレオチドを有する。

【0024】

本発明は更に、サンプル中の標的ポリヌクレオチドを検出する方法を提供し、該標的ポリヌクレオチドは(a)SEQ ID NO: 18-34を有する群から選択したポリヌクレオチド配列からなるポリヌクレオチド、(b)SEQ ID NO: 18-34を有する群から選択したポリヌクレオチド配列と少なくとも90%同一の天然のポリヌクレオチド配列からなるポリヌクレオチド、(c)(a)のポリヌクレオチドに相補的なポリヌクレオチド、(d)(b)のポリヌクレオチドに相補的なポリヌクレオチド、および(e)(a)~(d)のRNA等価物を含む群から選択されたポリヌクレオチドの配列を有する。検出方法は、(a)サンプル中の標的ポリヌクレオチドに相補的な配列からなる少なくとも20の連続したヌクレオチドを含むプローブを用いて該サンプルをハイブリダイズする過程と、(b)ハイブリダイゼーション複合体の存在・不存在を検出し、複合体が存在する場合にはオプシオンでその量を検出する過程からなり、プローブと標的ポリヌクレオチドあるいはその断片との間でハイブリダイゼーション複合体が形成されるような条件下で、プローブは標的ポリヌクレオチドに特異的にハイブリダイズする。一実施態様では、プローブは少なくとも60の連続したヌクレオチドを含む。

【0025】

本発明はまた、サンプル中の標的ポリヌクレオチドを検出する方法を提供する。ここで、標的ポリヌクレオチドは(a)SEQ ID NO: 18-34を有する群から選択したポリヌクレオチド配列からなるポリヌクレオチド、(b)SEQ ID NO: 18-34を有する群から選択したポリヌクレオチド配列と少なくとも90%同一である天然のポリヌクレオチド配列からなるポリヌクレオチド、(c)(a)のポリヌクレオチドに相補的なポリヌクレオチド、(d)(b)のポリヌクレオチドに相補的なポリヌクレオチド、または(e)(a)~(d)のRNA等価物を含む群から選択されたポリヌクレオチドの配列を有する。検出方法は、(a)ポリメラーゼ連鎖反応増幅を用いて標的ポリヌクレオチ

10

20

30

40

50

ドまたはその断片を増幅する過程と、(b) 標的ポリヌクレオチドまたはその断片の存在・不存在を検出し、該標的ポリヌクレオチドまたはその断片が存在する場合にはオプションでその量を検出する過程を含む。

【0026】

発明は更に、有効量のポリペプチドと薬剤として許容できる賦形剤とを含む成分を提供し、有効量のポリペプチドは、(a) SEQ ID NO: 1-17を有する群から選択したアミノ酸配列を有するポリペプチド、(b) SEQ ID NO: 1-17を有する群から選択したアミノ酸配列と少なくとも90%同一である天然のアミノ酸配列を有するポリペプチド、(c) SEQ ID NO: 1-17を有する群から選択したアミノ酸配列を有するポリペプチドの生物学的活性断片、または(d) SEQ ID NO: 1-17を有する群から選択したアミノ酸配列を有するポリペプチドの免疫抗原性断片を含む。一実施例では、SEQ ID NO: 1-17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GCRC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】

本発明はまた、(a) SEQ ID NO: 1-17を有する群から選択したアミノ酸配列を含むポリペプチド、(b) SEQ ID NO: 1-17を有する群から選択したアミノ酸配列と少なくとも90%の相同性を有する天然のアミノ酸配列を含むポリペプチド、(c) SEQ ID NO: 1-17を有する群から選択したアミノ酸配列のポリペプチドの生物学的活性断片、または(d) SEQ ID NO: 1-17を有する群から選択したアミノ酸配列のポリペプチドの免疫抗原性断片を含む群から選択されたアゴニストとしてのポリペプチドの有効性を確認するために化合物をスクリーニングする方法を提供する。スクリーニング方法は、(a) ポリペプチドを有するサンプルを化合物に曝す過程と、(b) サンプル中のアゴニスト活性を検出する過程とを含む。別法では、本発明は、この方法によって同定されたアゴニスト化合物と好適な医薬用賦形剤とを含む組成物を提供する。一実施態様では、本発明は機能性GCRCの発現低下に関連する疾患又は病状を治療する方法であって、そのような治療を必要とする患者に対して成分を投与する過程を含む方法を提供する。

20

【0028】

本発明は更に、(a) SEQ ID NO: 1-17を有する群から選択したアミノ酸配列を有するポリペプチド、(b) SEQ ID NO: 1-17を有する群から選択したアミノ酸配列と少なくとも90%が同一である天然のアミノ酸配列を有するポリペプチド、(c) SEQ ID NO: 1-17を有する群から選択したアミノ酸配列を有するポリペプチドの生物学的活性断片、または(d) SEQ ID NO: 1-17を有する群から選択したアミノ酸配列を有するポリペプチドの免疫抗原性断片を含むポリペプチドのアンタゴニストとしての有効性を確認するために化合物をスクリーニングする方法を提供する。スクリーニング方法は、(a) ポリペプチドを含むサンプルを化合物に曝す過程と、(b) サンプル中のアンタゴニスト活性を検出する過程とを含む。一実施態様で本発明は、この方法によって同定したアンタゴニスト化合物と薬剤として許容できる賦形剤とを含む成分を提供する。別の実施態様では、機能性GCRCの過剰発現に関連する疾患又は病状を治療する方法であって、そのような治療を必要とする患者に対して成分を投与する過程を含む方法を提供する。

30

40

【0029】

本発明は更に、(a) SEQ ID NO: 1-17を有する群から選択したアミノ酸配列を含むポリペプチド、(b) SEQ ID NO: 1-17を有する群から選択したアミノ酸配列と少なくとも90%の相同性を有する天然のアミノ酸配列を含むポリペプチド、(c) SEQ ID NO: 1-17を有する群から選択したアミノ酸配列を有するポリペプチドの生物学的活性断片、または(d) SEQ ID NO: 1-17を有する群から選択したアミノ酸配列を有するポリペプチドの免疫抗原性断片を含む群から選択したポリペプチドに特異結合する化合物をスクリーニングする方法を提供する。スクリーニング方法

50

は、(a)ポリペプチドを適切な条件下で少なくとも1つの試験化合物に結合させる過程と、(b)試験化合物とのポリペプチドの結合を検出し、それによってポリペプチドに特異結合する化合物を同定する過程とを含む。

【0030】

更に本発明は、(a)SEQ ID NO: 1-17からなる一群から選択されたアミノ酸配列からなるポリペプチドと、(b)SEQ ID NO: 1-17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からなるポリペプチドと、(c)SEQ ID NO: 1-17からなる一群から選択されたアミノ酸配列を有するポリペプチドの生物学的活性断片と、(d)SEQ ID NO: 1-17からなる一群から選択されたアミノ酸配列を有するポリペプチドの免疫原性断片とで構成される群から選択された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。

10

その方法は、(a)ポリペプチドの活性が許容された条件下で、ポリペプチドを少なくとも1つの試験化合物と混合させる過程と、(b)ポリペプチドの活性を試験化合物の存在下で算定する過程と、(c)試験化合物の存在下でのポリペプチドの活性を試験化合物の不存在下でのポリペプチドの活性と比較する過程とを含み、試験化合物の存在下でのポリペプチドの活性の変化は、ポリペプチドの活性を調節する化合物であることを意味する。

【0031】

本発明は更に、標的ポリヌクレオチドの変異発現の有効性を確認するために化合物をスクリーニングする方法を提供する。標的ポリヌクレオチドは、SEQ ID NO: 18-34を有する群から選択した配列を含む。スクリーニング方法は、(a)標的ポリヌクレオチドを含むサンプルを化合物に曝す過程と、(b)標的ポリヌクレオチドの変異発現を検出する過程とを含む。

20

【0032】

本発明は更に、(a)核酸を含む生物学的サンプルを試験化合物で処理する過程と、(b)(i)SEQ ID NO: 18-34を有する群から選択したポリヌクレオチド配列を含むポリヌクレオチド、(ii)SEQ ID NO: 18-34を有する群から選択したポリヌクレオチド配列と少なくとも90%の相同性を有する天然のポリヌクレオチド配列を含むポリヌクレオチド、(iii)(i)に相補的な配列を有するポリヌクレオチド、(iv)(ii)のポリヌクレオチドに相補的なポリヌクレオチド、(v)(i)~(iv)のRNA等価物を含む群から選択したポリヌクレオチドの少なくとも20の連続したヌクレオチドから構成されるプローブを用いて、処理した生物学的サンプルの核酸をハイブリダイズする過程とを含む試験化合物の毒性の算定方法を提供する。ハイブリダイゼーションは、上記プローブと生物学的サンプル中の標的ポリヌクレオチドの間に特定のハイブリダイゼーション複合体が形成されるような条件下で発生し、上記標的ポリヌクレオチドは、(i)SEQ ID NO: 18-34を有する群から選択したポリヌクレオチド配列を含むポリヌクレオチド、(ii)SEQ ID NO: 18-34を有する群から選択したポリヌクレオチド配列と少なくとも90%の相同性を有する天然のポリヌクレオチド配列を含むポリヌクレオチド、(iii)(i)のポリヌクレオチドに相補的な配列を有するポリヌクレオチド、(iv)(ii)のポリヌクレオチドに相補的なポリヌクレオチド、(v)(i)~(iv)のRNA等価物を含む群から選択する。或いは、標的ポリヌクレオチドは、上記(i)~(v)を含む群から選択したポリヌクレオチド配列の断片と、(c)ハイブリダイゼーション複合体の量を定量する過程と、(d)処理した生物学的サンプルのハイブリダイゼーション複合体の量を、非処理の生物学的サンプルのハイブリダイゼーション複合体の量と比較する過程を含み、処理した生物学的サンプルのハイブリダイゼーション複合体の量の差は、試験化合物の毒性を示す。

30

40

【0033】

(発明を実施するための形態)

本発明のタンパク質、ヌクレオチド配列及び方法について説明するが、その前に、説明した特定の装置、材料及び方法に本発明が限定されるものではなく、改変し得ることを理解

50

されたい。また、ここで使用する専門用語は特定の実施例を説明する目的で用いたものに過ぎず、特許請求の範囲にのみ限定される本発明の範囲を限定することを意図したのではないことも併せて理解されたい。

【0034】

請求の範囲及び明細書中で用いている単数形の「或る」及び「その(この)」の表記は、文脈から明らかにそうでないとされる場合を除いて複数ものを指す場合もあることに注意しなければならない。従って、例えば「或る宿主細胞」と記されている場合にはそのような宿主細胞が複数あることもあり、「或る抗体」と記されている場合には単数または複数の抗体、及び、当業者に公知の抗体の等価物等についても言及しているのである。

【0035】

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

【0036】

定義

用語「G C R E C」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G C R E Cのアミノ酸配列を指す。

【0037】

「アゴニスト」の語は、G C R E Cの生物学的活性を強化または擬態する分子を指す。アゴニストの例として、タンパク質、核酸、糖質、小分子その他の任意の化合物や成分を挙げることができるが、これらはG C R E Cと直接相互作用することによって、或いはG C R E Cが関与する生物学的経路の構成要素に作用することによって、G C R E Cの活性を調節する。

【0038】

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

【0039】

G C R E Cをコードする「変異(altered)」核酸配列は、種々のヌクレオチドを欠失、挿入または置換する核酸配列を有し、G C R E Cと同一またはG C R E Cの機能的特徴を少なくとも1つ有するポリペプチドを産出する。この定義に含まれるのは、G C R E C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或いは検出困難な多型現象と、G C R E Cをコードするポリヌクレオチド配列のための正常な染色体遺伝子座以外の遺伝子座に占めるような、対立遺伝子変異体に対する不適切或いは不測のハイブリダイゼーションである。コードされたタンパク質も「変異」し得るものであり、サイレント変化を生ぜしめて結果的に機能的に等価なG C R E Cとなるようなアミノ酸残基の欠失、挿入または置換を含み得る。計画的アミノ酸置換は、G C R E Cの生物学的または免疫学的活性が保持される限りにおいて、残基の極性、電荷、溶解度、疎水性、親水性、及び/または両親媒性特性の類似性にに基づき行い得る。例えば、

10

20

30

40

50

負に帯電したアミノ酸にはアスパラギン酸及びグルタミン酸があり、正に帯電したアミノ酸にはリジン及びアルギニンがある。親水性値が近似している非荷電極性側鎖を有するアミノ酸には、アスパラギンとグルタミン、セリンとスレオニンがある。親水性値が近似している非荷電側鎖を有するアミノ酸には、ロイシンとイソロイシンとバリン、グリシンとアラニン、フェニルアラニンとチロシンがある。

【0040】

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

10

【0041】

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

【0042】

「アンタゴニスト」の語は、GCRECの生物学的活性を阻害或いは弱める分子を指す。アンタゴニストとしては、抗体などのタンパク質、核酸、糖質、小分子またはその他の任意の化合物や成分を挙げることができるが、これらはGCRECと直接相互作用することによって、或いはGCRECが関与する生物学的経路の構成要素に作用することによって、GCRECの活性を調節する。

20

【0043】

「抗体」の語は、エピトープの決定基と結合することができる、そのままの免疫グロブリンやその断片、例えばF_a、F（a b'）₂及びF_v断片を指す。GCRECポリペプチドと結合する抗体は、免疫抗原として、そのままのポリペプチド、または関心のある小ペプチドを含む断片を用いて作製可能である。動物（マウス、ラット、ウサギ等）を免疫化するために用いるポリペプチドまたはオリゴペプチドは、RNAの翻訳、または化学合成によって得られるポリペプチドまたはオリゴペプチドに由来し得るもので、好みに応じて担体タンパク質に抱合することも可能である。通常用いられる担体であってペプチドと化学結合するものは、ウシ血清アルブミン、サイログロブリン及びスカイガイのヘモシアニン（KLH）等がある。結合ペプチドは、動物を免疫化するために用いる。

30

【0044】

「抗原決定基」の語は、特定の抗体と接触している分子の領域（即ちエピトープ）を指す。タンパク質またはタンパク質断片を用いて宿主動物を免疫化する場合、タンパク質の多数の領域が、抗原決定基（タンパク質の特定の領域または3次元構造）に特異結合する抗体の産生を誘導し得る。抗原決定基は、抗体への結合において無損傷抗原（即ち免疫応答を誘導するために用いられる免疫原）と競合し得る。

【0045】

「アンチセンス」の語は、特定の核酸配列の「センス」（コーディング）鎖と塩基対を形成することが可能な任意の成分を指す。アンチセンス成分には、DNAや、RNAや、ペプチド核酸（PNA）や、ホスホリボチオ酸、メチルホスホン酸またはベンジルホスホン酸等の修飾されたバックボーン連鎖を有するオリゴヌクレオチドや、2'-メトキシエチル糖または2'-メトキシエトキシ糖等の修飾された糖類を有するオリゴヌクレオチドや、或いは5-メチルシトシン、2-デオキシウラシルまたは7-デアザ-2'-デオキシグアノシン等の修飾された塩基を有するオリゴヌクレオチドが含まれる。アンチセンス分子は、化学合成または転写を含む任意の方法で製造することができる。相補的アンチセンス分子は、ひとたび細胞に導入されたら、細胞が形成した天然の核酸配列と塩基対を形成し、転写または翻訳を妨害する二重鎖を形成する。「負」または「マイナス」という表現は、ある参考DNA分子のアンチセンス鎖を意味し、「正」または「プラス」という表現は同センス鎖を意味することがある。

40

【0046】

50

「生物学的に活性」の語は、天然分子の構造的機能、調節機能または生化学的機能を有するタンパク質を指す。同様に「免疫学的に活性」はあるいは「免疫抗原性」は天然、組換えまたは合成の G C R E C、或いはその任意のオリゴペプチドの能力であって、適切な動物または細胞において特定の免疫反応を誘導して特定の抗体と結合し得る能力を指す。

【0047】

「相補(的)」または「相補性」の語は、塩基対形成によってアニーリングする2つの一本鎖核酸の間の関係を指す。例えば、配列「5' A - G - T 3'」は、相補配列「3' T - C - A 5'」と対を形成する。

【0048】

「所定のポリヌクレオチド配列を含む成分」及び「所定のアミノ酸配列を含む成分」は、所定のポリヌクレオチド配列またはアミノ酸配列を含む広範囲の任意の成分を指す。この成分には、乾燥製剤または水溶液が含まれ得る。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は、塩(例えばNaCl)、界面活性剤(例えばドデシル硫酸ナトリウム; SDS)及びその他の構成成分(例えばデンハート液、脱脂粉乳、サケの精子のDNA等)を含む水溶液中にプローブを分散させることができる。

10

【0049】

「コンセンサス配列」は、不要な塩基を分離するためにDNA配列の解析を繰り返し行い、XL-PCRキット(Applied 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。伸長及び構築の両方を行ってコンセンサス配列を決定する配列もある。

20

【0050】

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

30

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

10

20

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

【0051】

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

【0052】

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

30

【0053】

「検出可能な標識」は、測定可能な信号を発生することができ、ポリヌクレオチドまたはポリペプチドに共有結合または非共有結合するようなレポーター分子または酵素を指す。

【0054】

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

40

【0055】

「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部分であって、親配列と同一配列であるが親配列よりも長さが短い配列を指す。断片は、画定された配列の全長から1ヌクレオチド/アミノ酸残基を差し引いた長さよりも短い長さを有し得る。例えば或る断片は、5~1000の連続したヌクレオチドまたはアミノ酸残基を有し得る。プローブ、プライマー、抗原、治療用分子として、或いはその他の目的のために用いられる断片は、少なくとも5、10、15、16、20、25、30、40、50、

50

60、75、100、150、250若しくは500の連続したヌクレオチド或いはアミノ酸残基長さであり得る。断片は、分子の特定領域から優先的に選択し得る。例えば、ポリペプチド断片は、所定の配列に示すような最初の250または500アミノ酸（またはポリペプチドの最初の25%または50%）から選択された或る長さの連続したアミノ酸を有し得る。これらの長さは明らかに例として挙げているものであり、本発明の実施例では、配列表、表及び図面を含む明細書に裏付けされた任意の長さであってよい。

【0056】

SEQ ID NO: 18 - 34の断片には、固有のポリヌクレオチド配列領域が含まれる。この領域は、SEQ ID NO: 18 - 34を特異的に同定するものであり、例えばこの断片を得たゲノム中のSEQ ID NO: 18 - 34以外の配列とは異なるものである。SEQ ID NO: 18 - 34の断片は、例えば、ハイブリダイゼーション及び増幅技術において、或いは関連するポリヌクレオチド配列からSEQ ID NO: 18 - 34を区別する類似の方法において有用である。SEQ ID NO: 18 - 34の断片の正確な長さ及び断片に対応するSEQ ID NO: 18 - 34の領域は、断片に対する意図した目的に基づき当業者が慣例的に決定することが可能である。

10

【0057】

SEQ ID NO: 1 - 17の断片は、SEQ ID NO: 18 - 34の断片によってコードされる。SEQ ID NO: 1 - 17の断片には、SEQ ID NO: 1 - 17を特異的に同定する固有のアミノ酸配列領域が含まれている。例えば、SEQ ID NO: 1 - 17の断片は、SEQ ID NO: 1 - 17を特異認識する抗体を産出するための免疫抗原性ペプチドとして有用である。SEQ ID NO: 1 - 17の断片及び断片に対応するSEQ ID NO: 1 - 17の領域の正確な長さは、断片に対する意図した目的に基づき当業者が慣例的に決定することが可能である。

20

【0058】

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

【0059】

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

30

【0060】

ポリヌクレオチド配列に適用される「一致率」または「一致%」の語は、標準化されたアルゴリズムを用いてアラインメントされた少なくとも2つ以上のポリヌクレオチド配列間で一致する残基の割合を意味する。このようなアルゴリズムは、2配列間のアラインメントを最適化するために比較する配列において、標準化された再現性のある方法でギャップを挿入するので、2つの配列をより有意に比較できる。

【0061】

ポリヌクレオチド配列間の一一致率は、MEGALIGN version 3.12e配列アラインメントプログラムに組込まれているようなCLUSTAL Vアルゴリズムのデフォルトのパラメータを用いて決定できる。このプログラムは、LASERGENEソフトウェアパッケージ（一組の分子生物学的分析プログラム）（DNA STAR, Madison WI）の一部である。このCLUSTAL Vは、Higgins, D.G. 及びP.M. Sharp (1989) CABIOS 5: 151 - 153、Higgins, D.G. 他 (1992) CABIOS 8: 189 - 191. で説明されている。ポリヌクレオチド配列を2つ1組でアラインメントする際のデフォルトパラメータは、Ktuple = 2、gap penalty = 5、window = 4、「diagonals saved」= 4と設定する。「重み付けされた」残基重み付け表が、デフォルトとして選択された。CLUSTAL Vは、アラインメントされたポリヌクレオチド配列対間の「類似率」として一致率を報告する。

40

【0062】

50

或いは、米国国立バイオテクノロジー情報センター (NCBI) の Basic Local Alignment Search Tool (BLAST) が一般的に用いられ、且つ、無料で入手可能な配列比較アルゴリズム一式を提供している (Altschul, S. F. ら (1990) J. Mol. Biol. 215: 403-410)。このアルゴリズムは、幾つかの情報源から入手可能であり、メリーランド州ベセスダにある NCBI 及びインターネット (<http://www.ncbi.nlm.nih.gov/BLAST/>) から入手可能である。BLAST ソフトウェア一式には様々な配列分析プログラムが含まれており、既知のポリヌクレオチド配列を種々のデータベースから得た別のポリヌクレオチド配列とアラインメントする「blastn」もその1つである。その他にも、2つのヌクレオチド配列を対で直接比較するために用いる「BLAST 2 Sequences」と称されるツールも利用可能である。「BLAST 2 Sequences」は、<http://www.ncbi.nlm.nih.gov/gorf/bl2.html> にアクセスして対話形式で利用することが可能である。「BLAST 2 Sequences」ツールは、blastn と blastp (後述) の両方に用いることができる。BLAST プログラムは、一般的には、ギャップ及びデフォルト設定に設定された他のパラメータと共に用いる。例えば、2つのヌクレオチド配列を比較するために、デフォルトパラメータとして設定された「BLAST 2 Sequences」ツール Version 2.0.12 (2000年4月21日) を用いて blastn を実行してもよい。デフォルトパラメータの設定例を以下に示す。

10

20

【0063】

```
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
```

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

30

【0064】

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

40

【0065】

ポリペプチド配列に適用される「一致率」または「一致%」の語は、標準化されたアルゴリズムを用いてアラインメントされた少なくとも2以上のポリペプチド配列間で一致する残基の割合を意味する。ポリペプチド配列アラインメントの方法は公知である。保存的アミノ酸置換を考慮するアラインメント方法もある。既に詳述したこのような保存的置換は通常、置換部位の電荷及び疎水性を保存するので、ポリペプチドの構造を (従って機能も) 保存する。

【0066】

ポリペプチド配列間の一致率は、MEGALIGN version 3.12e 配列アラインメントプログラムに組込まれているような CLUSTAL V アルゴリズムのデフォ

50

ルトのパラメータを用いて決定できる（既に説明したのでそれを参照されたい）。CLUSTAL Vを用いて、ポリペプチド配列を2つ1組でアラインメントする際のデフォルトパラメータは、`Ktuple = 1`、`gap penalty = 3`、`window = 5`、`「diagonals saved」 = 5`と設定する。デフォルトの残基重み付け表としてPAM250マトリクスを選択する。ポリヌクレオチドアラインメントと同様に、CLUSTAL Vは、アラインメントされたポリペプチド配列対間の「類似率」として一致率を報告する。

【0067】

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

【0068】

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

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

【0069】

「ヒト人工染色体（HAC）」は、約6 kb（キロベース）～10 MbのサイズのDNA配列を含み得る、染色体の複製、分離及び維持に必要な全ての要素を含む直鎖状の小染色体である。

【0070】

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

【0071】

「ハイブリダイゼーション」は、所定のハイブリダイゼーション条件下で塩基対を形成することによって、一本鎖ポリヌクレオチドが相補的鎖とアニーリングするプロセスを指す。特異的ハイブリダイゼーションは、2つの核酸配列が高い相同性を共有することを示すものである。特異的ハイブリダイゼーション複合体は許容されるアニーリング条件下で形成され、「洗浄」ステップ後もハイブリダイズされたままである。洗浄ステップは、ハイブリダイゼーションプロセスのストリンジェンシーを決定する際に特に重要であり、更にストリンジェントな条件では、非特異結合（即ち完全には一致しない核酸鎖間の対の結合）が減少する。核酸配列のアニーリングに対する許容条件は、本技術分野における当業者が慣例的に決定する。許容条件はハイブリダイゼーション実験の間は一定でよいが、洗浄条件は所望のストリンジェンシーを得るように、従ってハイブリダイゼーション特異性も得るように実験中に変更することができる。アニーリングが許容される条件は、例えば、温度が68 で、約6 × SSC、約1% (w/v)のSDS、並びに約100 μg/mlのせん断して変性したサケ精子DNAが含まれる。

【0072】

10

20

30

40

50

一般に、ハイブリダイゼーションのストリンジェンシーは或る程度、洗浄ステップを実行する温度を基準にして表すことができる。このような洗浄温度は通常、所定のイオン強度及びpHにおける特定の配列の融点(T_m)より約5~20 低くなるように選択する。このT_mは、所定のイオン強度及びpHの下で、完全に一致するプローブに標的配列の50%がハイブリダイズする温度である。T_mを計算する式及び核酸のハイブリダイゼーション条件はよく知られており、Sambrookら(1989) Molecular Cloning: A Laboratory Manual, 第2版, 1-3巻, Cold Spring Harbor Press, Plainview NYに記載されており、特に2巻の9章を参照されたい。

【0073】

本発明のポリヌクレオチド間の高いストリンジェンシーのハイブリダイゼーションでは、約0.2×SSC及び約0.1%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、または42 の温度で行う。SSC濃度は、約0.1%のSDS存在下で、約0.1~2×SSCの範囲で変化し得る。通常は、ブロッキング剤を用いて非特異ハイブリダイゼーションを阻止する。このようなブロッキング剤には、例えば、約100~200 µg/mlのせん断して変性したサケ精子DNAがある。例えばRNAとDNAのハイブリダイゼーションのような特定条件下では、有機溶剤、例えば約35~50% v/vの濃度のホルムアミドを用いることもできる。洗浄条件の有用なバリエーションは、当業者には自明であろう。ハイブリダイゼーションは、特に高ストリンジェント条件下では、ヌクレオチド間の進化的な類似性を示唆し得る。このような類似性は、ヌクレオチド及びヌクレオチドにコードされるポリペプチドに対する類似の役割を強く示唆している。

10

20

【0074】

「ハイブリダイゼーション複合体」の語は、相補的塩基対間の水素結合の形成力によって2つの核酸配列間に形成された複合体を指す。ハイブリダイゼーション複合体は、溶解状態で形成し得る(C₀tまたはR₀t解析等)。或いは、一方の核酸配列が溶解状態で存在し、もう一方の核酸配列が固体支持体(例えば紙、膜、フィルタ、チップ、ピンまたはガラスライド、或いは他の適切な基質であって細胞若しくはその核酸が固定される基質)に固定されているような2つの核酸配列間に形成され得る。

【0075】

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

30

【0076】

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

【0077】

「免疫抗原性断片」とは、哺乳動物等の生命体に導入されると免疫応答を誘発し得るようなGCRCのポリペプチドまたはオリゴペプチド断片である。「免疫抗原性断片」の語には、本明細書中で開示したような或いは当分野で既知であるような任意の抗体産出方法において有用なGCRCの任意のポリペプチドまたはオリゴペプチド断片も含まれる。

40

【0078】

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

【0079】

「エレメント」または「アレイエレメント」の語は、マイクロアレイ上に定義された固有の位置にあるハイブリダイズ可能なポリヌクレオチド、ポリペプチドその他の化合物を指す。

【0080】

50

「調節(する)」の語は、G C R E Cの活性の変化を指す。調節することによって例えば、G C R E Cのタンパク質活性、結合特性その他の生物学的、機能的または免疫学的特性が増大または低下し得る。

【0081】

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

【0082】

「機能的に結合した」は、第1核酸配列が第2核酸配列と機能的な関係があるように配置された状態を指す。例えば、プロモーターがコード配列の転写または発現に影響を及ぼす場合には、そのプロモーターはそのコード配列に機能的に結合している。機能的に結合したDNA配列は非常に近接するか、或いは連続し得る。そして、2つのタンパク質コード領域を結合する必要がある場合は、同一のリーディングフレーム内にある。

10

【0083】

「ペプチド核酸」(PNA)は、アンチセンス分子または抗遺伝子物質であって、リジンを末端とするアミノ酸残基のペプチドバックボーンに結合した、少なくとも約5ヌクレオチドの長さのオリゴヌクレオチドからなるものを指す。末端のリジンは、成分に溶解性を与える。PNAは、相補的一本鎖DNAまたはRNAに優先的に結合して転写の拡張を停止するものであり、ポリエチレングリコール化して細胞におけるPNAの寿命を延長し得る。

20

【0084】

G C R E Cの「翻訳後修飾」は、脂質化、グリコシル化、リン酸化、アセチル化、ラセミ化、タンパク分解性開裂及びその他の当分野で既知の修飾を含み得る。これらのプロセスは、合成的または生化学的に発生し得る。生化学的修飾は、G C R E Cの酵素環境に依存し、細胞タイプによって変化し得る。

【0085】

「プローブ」は、G C R E C、G C R E Cの相補配列またはこれらの断片をコードする核酸配列を指し、同一核酸配列、対立遺伝子核酸配列または関連する核酸配列の検出に用いられる。プローブは、単離されたオリゴヌクレオチドまたはポリヌクレオチドであって、検出可能な標識またはレポーター分子に結合したものである。典型的な標識には、放射性アイソトープ、リガンド、化学発光試薬及び酵素がある。「プライマー」は、短い核酸、通常はDNAオリゴヌクレオチドであり、相補的塩基対を形成することで標的ポリヌクレオチドにアニーリングされ得る。プライマーは次に、DNAポリメラーゼ酵素によって標的DNA鎖に延在し得る。プライマー対は、例えばポリメラーゼ連鎖反応(PCR)による核酸配列の増幅(及び同定)に用い得る。

30

【0086】

本発明に用いるようなプローブ及びプライマーは通常、既知の配列の少なくとも15の連続したヌクレオチドを含んでいる。特異性を高めるために長めのプローブ及びプライマー、例えば開示した核酸配列の少なくとも20、25、30、40、50、60、70、80、90、100または150の連続したヌクレオチドからなるようなプローブ及びプライマーを用いてもよい。これよりもかなり長いプローブ及びプライマーもある。表、図面及び配列リストを含む本明細書に裏付けされた任意の長さのヌクレオチドを用いることができるものと理解されたい。

40

【0087】

プローブ及びプライマーの調製及び使用方法については、Sambrook, J. ら(1989) Molecular Cloning: A Laboratory Manual, 第2版, 1-3巻, Cold Spring Harbor Press, Plainview NY、Ausubel, F.M. ら, (1987) Curren

50

t Protocols in Molecular Biology, Greene Publi. Assoc. & Wiley-Intersciences, New York NY, Innisら (1990) PCR Protocols, A Guide to Methods and Applications Academic Press, San Diego CA等を参照されたい。PCRプライマー対は、その目的のためのコンピュータプログラム、例えばPrimer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA)を用いるなどして既知の配列から得ることができる。

【0088】

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

【0089】

「組換え核酸」は天然配列ではない配列であるか或いは人為的に組み合わせなければ離隔しているような配列の2以上のセグメントを人為的に組み合わせで産出した配列を有する配列である。この人為的組合せはしばしば化学合成によって達成するが、より一般的には核酸の単離セグメントの人為的操作によって、例えばのSambrookらの文献(前出)に記載されているような遺伝子工学的手法によって達成する。組換え核酸の語は、単に核酸の一部を付加、置換または欠失した変異核酸も含む。しばしば組換え核酸には、プロモーター配列に機能的に結合した核酸配列が含まれる。このような組換え核酸は、例えばある細胞を形質転換するために使用されるベクターの一部とすることが可能である。

【0090】

或いはこのような組換え核酸は、ウイルスベクターの不可欠な要素であって例えばワクシニアウイルスに基づくものであり得る。ワクシニアウイルスは哺乳動物のワクチン接種に用いることが可能で、その際に組換え核酸が発現して哺乳動物の防御免疫応答を誘導する。

【0091】

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

【0092】

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

【0093】

DNA配列に関する「RNA等価物」は、窒素塩基チミンが全てウラシルに置換されていることと、糖のバックボーンがデオキシリボースではなくリボースから構成されていることを除いて、参照DNA配列と同一のヌクレオチド線形配列から構成されている。

【0094】

「サンプル」の語は、その最も広い意味で用いられる。GCRCをコードする核酸若しくはその断片、またはGCRC自体を含む疑いのあるサンプルは、体液や、細胞、染色体、細胞小器官または細胞から単離された膜からの抽出物や、細胞や、溶解しているか基質に結合しているゲノムDNA、RNAまたはcDNAや、組織や、組織プリント等から構成され得る。

【0095】

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

【0096】

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

【0097】

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

【0098】

「基質」は、任意の好適な固体または半固体の支持体を指すものであって、膜、フィルタ、チップ、スライド、ウエハ、ファイバー、磁性または非磁性ビーズ、ゲル、管、プレート、ポリマー、微細粒子、毛管が含まれる。基質は、凹み、溝、ピン、チャネル、孔等、様々な表面形態を有することができ、基質表面にはポリヌクレオチドやポリペプチドが結合する。

【0099】

「転写イメージ」は、所与の時間、条件での固有の細胞タイプまたは組織による遺伝子発現の集合的パターンを指す。

【0100】

「形質転換」は、外来性のDNAが受入細胞に入り込むプロセスを表す。形質転換は、本技術分野で知られている種々の方法に従って自然条件または人工条件下で生じ得るものであり、外来性の核酸配列を原核または真核宿主細胞に挿入する任意の既知の方法を基にし得る。形質転換の方法は、形質転換する宿主細胞の種類によって選択する。限定するものではないが形質転換方法には、バクテリオファージまたはウイルス感染、電気穿孔法(

10

20

30

40

50

エレクトロポレーション)、熱ショック、リポフェクション及び微粒子銃を用いる方法がある。「形質転換された」細胞には、導入されたDNAが自律的に複製するプラスミドとして或いは宿主染色体の一部として複製可能である安定的に形質転換された細胞が含まれる。さらに、限られた時間に一時的に導入DNA若しくは導入RNAを発現する細胞も含まれる。

【0101】

ここで用いる「遺伝形質転換体」とは任意の有機体であり、限定するものではないが動植物を含み、有機体の1個若しくは数個の細胞が、ヒトの関与によって、例えば本技術分野でよく知られている形質転換技術によって導入された異種核酸を有する。核酸の細胞への導入は、直接または間接的に、細胞の前駆物質に導入することによって、計画的な遺伝子操作によって、例えば微量注射法によって或いは組換えウイルスの感染によって行う。遺伝子操作の語は、古典的な交雑育種或いは*in vitro*受精を指すものではなく、組換えDNA分子の導入を指すものである。本発明に基づいて予期される遺伝形質転換体には、バクテリア、シアノバクテリア、真菌及び動植物がある。本発明の単離されたDNAは、本技術分野で知られている方法、例えば感染、形質移入、形質転換またはトランス接合によって宿主に導入することができる。本発明のDNAをこのような有機体に移入する技術はよく知られており、前述のSambrookら。(1989)等の参考文献に記載されている。

10

【0102】

特定の核酸配列の「変異体」は、核酸配列1本全部の長さに対して特定の核酸配列と少なくとも40%の相同性を有する核酸配列であると定義する。その際、デフォルトパラメータに設定した「BLAST 2 Sequences」ツールVersion 2.0.9(1999年5月7日)を用いてblastnを実行する。このような核酸対は、所定の長さに対して、例えば少なくとも50%、60%、70%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%またはそれ以上の相同性を示し得る。或る変異体は、例えば「対立遺伝子」変異体(前述)、「スプライス」変異体、「種」変異体または「多形性」変異体として説明し得る。スプライス変異体は参照分子とかなりの相同性を有し得るが、mRNAプロセッシング中のエキソンの交互スプライシングによって通常多数の或いは僅かな数のポリヌクレオチドを有することになる。対応するポリペプチドは、追加機能ドメインを有するか或いは参照分子に存在するドメインが欠落していることがある。種変異体は、種相互に異なるポリヌクレオチド配列である。結果的に生じるポリペプチドは通常、相互にかなりのアミノ酸相同性を有する。多形性変異体は、与えられた種の個体間で特定の遺伝子のポリヌクレオチド配列が異なる。また、多形性変異体は、1つのヌクレオチド塩基によってポリヌクレオチド配列が変化する「1塩基性多形性」(SNP)を含み得る。SNPの存在は、例えば特定の個体群、病状または病状性向を示し得る。

20

30

【0103】

特定のポリペプチド配列の「変異体」は、ポリペプチド配列の1本の長さ全体で特定のポリペプチド配列に対して少なくとも40%の相同性を有するポリペプチド配列として画定される。ここで、デフォルトパラメータに設定した「BLAST 2 Sequence」ツールVersion 2.0.9(1999年5月7日)を用いてblastpを実行する。このようなポリペプチド対は、所定の長さに対して、例えば少なくとも50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%またはそれ以上の配列同一性を示し得る。

40

【0104】

発明

本発明は、新規なヒトGタンパク質(GCREC)と、GCRECをコードするポリヌクレオチドと、細胞増殖異常、神経障害、心血管障害、胃腸障害、自己免疫/炎症疾患及び代謝障害並びにウイルス性感染症の診断、治療並びに予防にこれらの配列を利用する方法の発見に基づくものである。

50

【0105】

表1は、本発明の完全長ポリヌクレオチド配列及びポリペプチド配列の命名の概略である。各ポリヌクレオチド及びその対応するポリペプチドは、1つのIncyteプロジェクト識別番号(IncyteプロジェクトID)と関連する。各ポリペプチド配列は、ポリペプチド配列識別番号(ポリペプチドSEQ ID NO)とIncyteポリペプチド配列番号(IncyteポリペプチドID)によって表示した。各ポリヌクレオチド配列は、ポリヌクレオチド配列識別番号(ポリヌクレオチドSEQ ID NO)とIncyteポリヌクレオチドコンセンサス配列番号(IncyteポリヌクレオチドID)によって表示した。

【0106】

表2は、GenBankタンパク質(genpept)データベースに対するBLAST分析によって同定されたような、本発明のポリペプチドとの相同性を有する配列を示している。列1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Incyte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、GenBank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。

【0107】

表3は、本発明のポリペプチドの様々な構造的特徴を示す。列1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Incyte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、リン酸化およびグリコシル化の可能性のある部位を示す。列6は、シグネチャ(signature)配列、ドメイン、およびモチーフを含むアミノ酸残基を示す。列7は、タンパク質の構造/機能の分析のための分析方法を示し、該当箇所にはさらに分析方法に利用した検索可能なデータベースを示す。

【0108】

合わせて、表2及び表3は本発明のポリペプチドの特性を概略的に示している。これらの特性は、特許請求の範囲に記載されたポリペプチドがGタンパク質結合受容体であることを証明する。例えば、SEQ ID NO: 1はヒトのオルファンGタンパク質結合受容体、GPC-R(GenBank ID g2865470)と99%の同一性を有することがBasic Local Alignment Search Tool (BLAST)によって示された(表2参照)。BLAST確率スコアは $1.5e^{-217}$ であり、これは観測されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO: 1はまた、オルファンGタンパク質結合受容体ドメインを有する。これは、遺伝子ファミリー、配列相同性および構造的フィンガープリント領域を検索するLocks IMProved Searcher (BLIMPS)を利用してDOMOデータベースにおいて、統計的に有意な一致を検索して決定された(表3参照)。SEQ ID NO: 1はまた、Gタンパク質結合受容体サブクラス膜7回貫通受容体(ロドプシンファミリー)ドメインを有する。これは、隠れマルコフモデル(HMM)を基にした保存されたタンパク質ファミリードメインのPFAMデータベースにおいて、統計的に有意な一致を検索して決定された(表3参照)。BLIMPS-PRINTS、BLIMPS-BLOCKS、MOTIFS及びPROFILESCAN分析から得たデータは、SEQ ID NO: 1がGタンパク質結合受容体であることを裏づける実証的な証拠を更に提供する。別の例においてSEQ ID NO: 8は、BLASTにより決定されるラッ

10

20

30

40

50

トセロトニン受容体 S 1 9 (GenBank ID g 3 1 0 0 7 5) に 7 9 % の同一性を有する。(表 2 参照) BLAST 確率スコアは 2.0×10^{-151} である。SEQ ID NO: 8 はまた、G タンパク質結合受容体ドメイン構造を有する。これは、隠れマルコフモデル (HMM) を基にした保存されたタンパク質ファミリードメインの PFAM データベースにおいて、統計的に有意な一致を検索して決定された (表 3 参照)。PRODOM 及び DOMO データベースのタンパク質シグネチャ配列への BLAST 比較そして BLIMPS 及び PROFILESCAN 分析から得たデータは、配列番号 8 が G タンパク質結合受容体であることを裏づける証拠を更に提供する。別の例において、SEQ ID NO: 1 0 は BLAST によって同定されるヒトの白血球血小板活性因子受容体、G タンパク質結合受容体 (GenBank ID g 1 8 9 2 7 0) と 9 2 % の同一性を有する (表 2 参照)。BLAST 確率スコアは 2.9×10^{-147} である。SEQ ID NO: 1 0 はまた受容体結合ドメインそして膜 7 回貫通受容体ドメイン、G タンパク質結合受容体の特徴を有する。これは、隠れマルコフモデル (HMM) を基にした保存されたタンパク質ファミリードメインの PFAM データベースにおいて、統計的に有意な一致を検索して決定された (表 3 参照)。BLIMPS、MOTIFS 分析から得たデータは、SEQ ID NO: 1 0 が G タンパク質結合受容体であることを裏づける証拠を更に提供する。別の例において SEQ ID NO: 1 1 は、BLAST により決定されるヒトのエンドセリン受容体 B 3 (GenBank ID g 4 5 8 0 9 2 4) に 9 9 % の同一性を有する。(表 2 参照) BLAST 確率スコアは 2.3×10^{-289} である。SEQ ID NO: 1 1 はまた、膜 7 回貫通受容体ドメインを有する。これは、隠れマルコフモデル (HMM) を基にした保存されたタンパク質ファミリードメインの PFAM データベースにおいて、統計的に有意な一致を検索して決定された。(表 3 参照) BLIMPS、MOTIFS、及び PROFILESCAN 解析よりのデータは、SEQ ID NO: 1 1 がエンドセリン受容体である、さらに実証的な証拠を提供する。別の例において SEQ ID NO: 1 6 は、BLAST により決定されるラット味蕾受容体タンパク質 TB 6 4 1 (GenBank ID g 1 2 5 6 3 9 3) に 5 0 % の同一性を有する (表 2 参照)。BLAST 確率スコアは 2.0×10^{-59} である。さらに PRODOM 及び DOMO データベースの BLAST 検索は SEQ ID NO: 1 6 が受容体タンパク質また G タンパク質結合受容体ドメインと相同性を有することも示す。SEQ ID NO: 1 6 はまた、ロドプシンファミリー膜 7 回貫通受容体ドメインを有するが、これは、隠れマルコフモデル (HMM) を基にした保存されたタンパク質ファミリードメインの PFAM データベースにおいて、統計的に有意な一致を検索して決定された (表 3 参照)。BLIMPS、PROFILESCAN 分析から得たデータは、SEQ ID NO: 1 6 が G タンパク質結合受容体であることを裏づける証拠を更に提供する。別の例において SEQ ID NO: 1 7 は、BLAST により決定されるラット G タンパク質結合受容体 OL 1 (GenBank ID g 1 0 1 6 3 6 2) に 4 1 % の同一性を有する。(表 2 参照) BLAST 確率スコアは 3.6×10^{-69} である。SEQ ID NO: 1 7 はまた、膜 7 回貫通受容体 (ロドプシンファミリー) ドメインを有する。これは、隠れマルコフモデル (HMM) を基にした保存されたタンパク質ファミリードメインの PFAM データベースにおいて、統計的に有意な一致を検索して決定された (表 3 参照)。BLIMPS、MOTIFS 及び PROFILESCAN 分析から得たデータは、SEQ ID NO: 1 7 がロドプシンファミリーの G タンパク質結合受容体であることを裏づける証拠を更に提供する。SEQ ID NO: 2 - 7、SEQ ID NO: 9 および SEQ ID NO: 1 2 - 1 5 は同じような方法で解析し、注釈を付けた。SEQ ID NO: 1 - 1 7 の解析のためのアルゴリズム及びパラメータが表 7 で記述されている。

【 0 1 0 9 】

表 4 に示すように、本発明の完全長ポリヌクレオチド配列は、cDNA 配列またはゲノム DNA 由来のコード (エキソン) 配列を用いて、或いはこれら 2 種類の配列を任意に組み合わせ構築した。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

10

20

30

40

50

れに対応する *Incyte* ポリヌクレオチドコンセンサス配列番号 (*Incyte Poly nucleotide ID*) を示す。列 3 は、各ポリヌクレオチド配列の長さ (塩基対単位) を示す。列 4 は、例えば、SEQ ID NO: 18 - 34 を同定するため、或いは SEQ ID NO: 18 - 34 と関連するポリヌクレオチド配列とを区別するためのハイブリダイゼーションまたは増幅技術に有用なポリヌクレオチド配列の断片を示す。列 5 は cDNA 配列、ゲノム DNA から予想されたコード配列 (エキソン) 及び/または cDNA 及びゲノム DNA を共に有する配列集合に対応する識別番号を示している。これらの配列は、本発明の完全長ポリヌクレオチド配列をアセンブル構築するのに用いた。表 4 の列 6 および列 7 はそれぞれ完全長配列に対して、列 5 の配列に対応する cDNA 配列およびゲノム配列の開始ヌクレオチド (5') 位置および終了ヌクレオチド (3') 位置を示す。 10

【0110】

表 4 の列 5 の識別番号は、特に例えば *Incyte* cDNA とそれに対応する cDNA ライブラリを指す場合もある。例えば、2432516H1 は *Incyte* cDNA 配列の識別番号であり、BRAVUNT02 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 *Incyte* cDNA は、プールされている cDNA ライブラリ (例えば、71687857V1) に由来する。或いは列 5 の識別番号は、完全長ポリヌクレオチド配列のアセンブリに寄与する GenBank cDNA または EST を指す場合もある。或いは列 5 の識別番号は、ゲノム DNA の Genscan 分析により予想されるコード領域を指す場合もある。例えば、GNN.g6479070.edit は、Genscan 推定コード配列の識別番号であって、g6479070 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、「エキソンステッチング (exon-stitching)」アルゴリズムにより結び合わせた cDNA 及び Genscan 予想エキソンの両方からなる群を指す場合もある。例えば、FL210313__00001 は、210313 がアルゴリズムが採用された配列のクラスタの識別番号であり、00001 がアルゴリズムによって生成された予測数である、ステッチ配列を示している (実施例 5 を参照)。または列 5 の識別番号は、「エキソンストレッチング (exon-stretching)」アルゴリズムにより結び合わせた cDNA 及び Genscan 予想エキソンの 30 両方からなる群を指す場合がある。例えば、FL7655614__g7406476__g156725 は「ストレッチされた」配列の識別番号である。7655614 は *Incyte* プロジェクト識別番号であり、g7406476 は「エキソンストレッチング」アルゴリズムが適用されたヒトゲノム配列の GenBank の識別番号であり、また g156725 は最も近い GenBank タンパク質相同体の GenBank の識別番号である。(実施例 5 を参照)。場合によっては、列 5 に示されている配列の範囲と重複する *Incyte* cDNA の範囲が得られ、最終的なコンセンサスポリヌクレオチド配列が決定されるが、それに相当する *Incyte* cDNA の識別番号は示されていない。

【0111】

表 5 は、*Incyte* cDNA 配列を用いて構築された完全長ポリヌクレオチド配列のための代表的な cDNA ライブラリを示している。代表的な cDNA ライブラリは、上記のポリヌクレオチド配列を構築及び確認するために用いられる *Incyte* cDNA 配列によって最も頻繁に表される *Incyte* cDNA ライブラリである。cDNA ライブラリを作製するために用いた組織及びベクターを表 5 に示し、表 6 で説明している。 40

【0112】

表 8 は、本発明のポリヌクレオチドの組織特異発現を示す。列 1 は、ポリヌクレオチドの発現のためにポリメラーゼ連鎖反応 (PCR) で試験された組織群を示す。残りの列は、特定のポリヌクレオチドが各組織群で発現したかどうかを示す。PCR 産物が検出された場合、“+” の記号で表示されるポジティブな発現により示した。また PCR 産物が検出できなかった場合“-” の記号で表示され、発現がなかったことを示した。 50

【0113】

本発明には、G C R E Cの変異体も含まれる。好適なG C R E Cの変異体のアミノ酸配列は、G C R E Cアミノ酸配列と少なくとも約80%、または少なくとも約90%、或いは少なくとも約95%もの一致率を有し、G C R E Cの機能的若しくは構造的特徴を少なくとも1つ有するような変異体である。

【0114】

本発明には、G C R E Cをコードするポリヌクレオチドも含まれる。或る例では、G C R E CをコードするS E Q I D N O : 1 8 - 3 4からなる群から選択された配列を有するポリヌクレオチド配列が本発明に含まれている。S E Q I D N O : 1 8 - 3 4のポリヌクレオチド配列は、配列表に示されているように等価R N A配列と同等の価値を有しているが、窒素塩基チミンの出現はウラシルに置換され、糖のバックボーンはデオキシリボースではなくシリボースから構成されている。

10

【0115】

本発明には、G C R E Cをコードするポリヌクレオチド配列の変異配列も含まれる。詳細には、このようなポリヌクレオチド配列の変異配列は、G C R E Cをコードするポリヌクレオチド配列に対して少なくとも70%のポリヌクレオチド配列同一性、或いは少なくとも85%のポリヌクレオチド配列同一性、更には少なくとも95%ものポリヌクレオチド配列同一性を有する。本発明の特定の実施形態は、S E Q I D N O : 1 8 - 3 4からなる一群から選択された核酸配列と少なくとも70%のポリヌクレオチド配列同一性、或いは少なくとも85%のポリヌクレオチド配列同一性、更には少なくとも95%ものポリヌクレオチド配列同一性を有するS E Q I D N O : 1 8 - 3 4からなる一群から選択された配列を含むポリヌクレオチド配列の変異配列を提供する。上記したポリヌクレオチド変異配列は何れも、G C R E Cの機能的或いは構造的特徴の少なくとも1つを有するアミノ酸配列をコードする。

20

【0116】

当業者であれば、遺伝暗号の縮重の結果、G C R E Cをコードする多数のポリヌクレオチド配列（既知の遺伝子または天然の遺伝子のポリヌクレオチド配列と最低限の類似性しか有しないポリヌクレオチド配列もある）を産出し得ることは理解できよう。したがって本発明には、可能コドン選択に基づく組合せの選択によって産出し得るようなありとあらゆる可能性のあるポリヌクレオチド配列変異体を網羅し得る。これらの組合せは、天然のG C R E Cのポリヌクレオチド配列に適用されるような標準トリプレット遺伝暗号を基に作られるものであり、このような変異は全て明確に開示されているものと考えられる。

30

【0117】

G C R E C及びその変異体をコードするヌクレオチド配列は通常、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、G C R E Cまたはその誘導体であって実質上異なるコドンの使用法があるもの、例えば天然に存在しないコドンの封入があるものをコードするヌクレオチド配列を作り出すことは有益であろう。宿主が特定のコドンを利用する頻度に基づいて、特定の真核又は原核宿主に発生するペプチドの発現率を高めるようにコドンを選択することが可能である。コードされたアミノ酸配列を変えることなく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上変更する別の理由には、天然の配列から作製される転写物より好ましい、例えば半減期が長いなどの特性を有するR N A転写物の作製がある。

40

【0118】

本発明には、G C R E C、G C R E C誘導体及びこれらの断片をコードするD N A配列又はそれらの断片を完全に合成化学によって作製することも含まれる。作製後、当分野でよく知られている試薬を用いて、この合成配列を任意の様々な入手可能な発現ベクター及び細胞系中に挿入し得る。更に、合成化学を用いてG C R E Cまたはその任意の断片をコードする配列に突然変異を誘導し得る。

【0119】

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

50

チド配列、特に、SEQ ID NO: 18 - 34 及びそれらの断片とハイブリダイズ可能なポリヌクレオチド配列が含まれる (例えば、Wahl, G. M. 及び S. L. Berger (1987) *Methods Enzymol.* 152: 399 - 407; Kimmel, A. R. (1987) *Methods Enzymol.* 152: 507 - 511.) アニール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。

【0120】

DNAシーケンシングの方法は当分野でよく知られており、本発明の何れの実施例もDNAシーケンシング方法を用いて実施可能である。DNAシーケンシング方法には酵素を用いることができ、例えばDNAポリメラーゼIのクレノウ断片、SEQUENASE (US Biochemical, Cleveland OH)、Taqポリメラーゼ (Applied Biosystems)、熱安定性T7ポリメラーゼ (Amersham, Pharmacia Biotech, Piscataway NJ) を用いることができる。或いは、例えばELONGASE増幅システム (Life Technologies, Gaithersburg MD) において見られるように、ポリメラーゼと校正エキソヌクレアーゼを併用することができる。好適には、MICROLAB 2200液体転移システム (Hamilton, Reno, NV)、PTC 200サーマルサイクラー (MJ Research, Watertown MA) 及びABI CATALYST 800サーマルサイクラー (Applied Biosystems) 等の装置を用いて配列の調製を自動化する。次に、ABI 373 或いは 377 DNAシーケンシングシステム (Applied Biosystems)、MEGABACE 1000 DNAシーケンシングシステム (Molecular Dynamics, Sunnyvale CA) または当分野でよく知られている他の方法を用いてシーケンシングを行う。結果として得られた配列を当分野でよく知られている種々のアルゴリズムを用いて分析する (Ausubel, F. M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7、Meyers, R. A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, 856 - 853ページ等を参照)。

【0121】

GC RECをコードする核酸配列を、部分的ヌクレオチド配列を利用し且つ当分野でよく知られているPCR法をベースにした種々の方法を用いて伸長させ、プロモーター及び調節要素等の上流配列を検出することができる。例えば、使用し得る方法の一つである制限部位PCR法は、ユニバーサルプライマー及びネステッドプライマーを用いてクローニングベクター内のゲノムDNAから未知の配列を増幅する方法である (Sarkar, G. (1993) *PCR Methods Applic.* 2: 318322. を参照)。別の方法に逆PCR法があり、これは広範な方向に伸長させたプライマーを用いて環状化した鋳型から未知の配列を増幅する方法である。鋳型は、既知のゲノム遺伝子座及びその周辺の配列を含む制限酵素断片から得る (Triglia, T. 他 (1988) *Nucleic Acids Res.* 16: 8186. を参照)。第3の方法としてキャプチャPCR法があり、これはヒト及び酵母菌人工染色体DNAの既知の配列に隣接するDNA断片をPCR増幅する方法に参与している (Lagerstrom, M. 他 (1991) *PCR Methods Applic.* 1: 111119. を参照)。この方法では、PCRを行う前に複数の制限酵素の消化及び連結反応を用いて未知の配列領域内に組換え二本鎖配列を挿入することが可能である。また、未知の配列を検索するために用い得る別の方法については当分野で知られている (Parker, J. D. ら (1991) *Nucleic Acids Res.* 19: 30553060等を参照)。更に、PCR、ネステッドプライマー及びPromoter Finder (商標) ライブラリ (Clontech, Palo Alto CA) を用いてゲノムDNAをウォーキングすることができる。この手順は、ライブラリをスクリーニングする必要がなく、イントロンノ

エキソン接合部を見付けるのに有用である。全てのPCRベースの方法に対して、市販されているソフトウェア、例えばOLIGO 4.06プライマー分析ソフトウェア(National Biosciences, Plymouth MN)或いは別の好適なプログラムを用いて、長さが約22~30ヌクレオチド、GC含有率が約50%以上、温度約68~72で鋳型に対してアニーリングするようにプライマーを設計し得る。

【0122】

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

10

【0123】

市販されているキャピラリー電気泳動システムを用いて、シーケンシングまたはPCR産物のサイズを分析し、またはそのヌクレオチド配列を確認することができる。具体的には、キャピラリーシーケンシングは、電気泳動による分離のための流動性ポリマーと、4つの異なるヌクレオチドに特異的であるような、レーザで活性化される蛍光色素と、放出された波長の検出に利用するCCDカメラとを有し得る。出力/光の強度は、適切なソフトウェア(Applied Biosystems社のGENOTYPER、SEQUENCE NAVIGATOR等)を用いて電気信号に変換し得る。サンプルのロードからコンピュータ分析及び電子データ表示までの全プロセスがコンピュータ制御可能である。

20

【0124】

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、適切な宿主細胞内でGCREC、GCRECの断片またはその機能的等価物を発現させるような組換えDNA分子内でクローニングし得る。遺伝暗号固有の宿重に起因して、実質的同一或いは機能的等価のアミノ酸配列をコードするような別のDNA配列を作製してGCRECの発現に利用し得る。

【0125】

種々の目的でGCRECがコードする配列を変えるために、本発明のヌクレオチド配列を当分野で通常知られている方法を用いて組み換えることができる。ここで目的には、限定するものではないが遺伝子産物のクローニング、プロセッシング、発現の調節がある。遺伝子断片及び合成オリゴヌクレオチドのランダムなフラグメンテーション及びPCR再アセンブリによるDNAシャッフリングを用い、ヌクレオチド配列を組み換えることが可能である。例えば、オリゴヌクレオチド仲介特定部位突然変異誘導を利用して、新規な制限部位の生成、グリコシル化パターンの変更、コドン優先の変更、スプライス変異体の生成等を行う突然変異を導入し得る。

30

【0126】

本発明のヌクレオチドは、MOLECULAR BREEDING(Maxygen Inc., Santa Clara CA. 米国特許第5,837,458号、Chang, C.-C.ら(1999) Nat. Biotechnol. 17:793-797、Christians, F.C.ら(1999) Nat. Biotechnol. 17:259-264、Cramer, A.ら(1996) Nat. Biotechnol. 14:315-319に記載)等のDNAシャッフリング技術の対象となり、GCRECの生物学的特性、例えば生物活性、酵素力、或いは他の分子や化合物との結合力等を変更または向上させ得る。DNAシャッフリングは、遺伝子断片のPCR仲介再組換えを用いて遺伝子変異体のライブラリを生成するプロセスである。ライブラリはその後、その遺伝子変異体を所望の特性に同定するような選択またはスクリーニングにかけられる。次にこれらの好適な変異体をプールし、更に反復してDNAシャッフリング及び選択/スクリーニングを行ってもよい。このように、遺伝の多様性は「人為的」品種改良及び

40

50

急速な分子の進化を経て創生される。例えば、ランダムポイント突然変異を有する単一の遺伝子の断片を組み換えて、スクリーニングし、その後所望の特性が最適化されるまでシャッフリングすることができる。或いは、所定の遺伝子を同種または異種のいずれかから得た同一遺伝子ファミリーの相同遺伝子と組み換えて、それによって天然に存在する複数の遺伝子の遺伝多様性を、指図された制御可能な方法で最大化させることができる。

【0127】

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. 7: 225-232*を参照)。或いは、化学的方法を用いてGCRECそれ自体またはその断片を合成し得る。例えば、種々の液相または固相技術を用いてペプチド合成を行うことができる(Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, 55-60ページ、Roberge, J.Y.ら(1995) *Science* 269: 202204等を参照)。自動合成はABI 431Aペプチドシンセサイザ(Applied Biosystems)を用いて達成し得る。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。

【0128】

ペプチドは、分離用高速液体クロマトグラフィーを用いて実質上精製可能である(Chiez, R.M.及びF.Z. Regnier (1990) *Methods Enzymol.* 182: 392-421等を参照)。合成ペプチドの組成は、アミノ酸分析またはシーケンシングによって確認することができる(前出のCreighton, 28-53ページ等を参照)。

【0129】

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とができる。好適な発現ベクターとは即ち好適な宿主内で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要素を含むベクターである。必要な要素には、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プロモーター、5'及び3'の非翻訳領域などの調節配列がある。このような要素は、強度及び特異性が様々である。固有開始シグナルを用いて、GCRECをコードする配列をより効果的に翻訳することが可能もある。このようなシグナルには、ATG開始コドンと、コザック配列などの近傍の配列が含まれる。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るようにすべきである。外来性の翻訳要素及び開始コドンは、様々な天然物及び合成物を起源とし得る。用いられる特定の宿主細胞系に好適なエンハンサーを含めることで発現の効率を高めることが可能である(Scharf, D.ら(1994) *Results Probl. Cell Differ.* 20: 125-162.等を参照)。

【0130】

当業者によく知られている方法を用いて、GCRECをコードする配列と、好適な転写及び翻訳調節要素とを含む発現ベクターを構築することが可能である。これらの方法には、in vitro組換えDNA技術、合成技術、及びin vivo遺伝子組換え技術が含まれる(例えば、Sambrook, J.他.(1989) Molecular C

loning. A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, 4章、8章及び16-17章、Ausubel, F.M. 他. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY、9章、13章及び16章を参照)。

【0131】

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

【0132】

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、サブクローニング及び増殖には、PBLUESCRIPT(Stratagene, La Jolla CA)またはPSPORT1プラスミド(Life Technologies)などの多機能大腸菌ベクターを用いることができる。GCRECをコードする配列の、ベクターの多数のクローニング部位への連結反応によって、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を同定するための比色スクリーニング法が可能となる。更にこれらのベクターは、クローニングされた配列における*in vitro*転写、ジデオキシのシークエンシング、ヘルパーファージによる一本鎖の救出、入れ子状態の欠失の生成にも有用であろう(例えば、Van Heeke, G.およびS.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509を参照)。多量のGCRECが必要な場合、例えば抗体を生成する場合などには、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、強力な誘導SP6バクテリオファージプロモーターまた

は誘導T7バクテリオファージプロモーターを含むベクターが使用できる。

【0133】

酵母の発現系を使用してGCRECを生成し得る。因子、アルコールオキシダーゼ、PGHプロモーター等の構成型或いは誘導型のプロモーターを含む多数のベクターが、出芽酵母菌 (*Saccharomyces cerevisiae*) またはピキア酵母 (*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)。

【0134】

植物系を使用してGCRECを発現することも可能である。GCRECをコードする配列の転写は、ウイルスプロモーター、例えば単独或いはTMV (*Takamatsu, N. (1987) EMBO J* 6: 307-311) 由来のオメガリーダー配列と組み合わせ用いられるようなCaMV由来の35S及び19Sプロモーターによって促進される。或いは、RUBISCOの小サブユニット等の植物プロモーターまたは熱ショックプロモーターを用いてもよい (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) 。これらの構成物は、直接DNA形質転換または病原体を媒介とする形質移入によって、植物細胞内に導入可能である (『マグローヒル科学技術年鑑』 (*The McGraw Hill Yearbook of Science and Technology*) (1992) McGraw Hill New York NY, 191-196 ページ等を参照)。

【0135】

哺乳動物細胞においては、多数のウイルスベースの発現系を利用し得る。発現ベクターとしてアデノウイルスを用いる場合、GCRECをコードする配列は、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/翻訳複合体に連結反応され得る。非必須のE1またはE3領域へウイルスのゲノムを挿入し、宿主細胞でGCRECを発現する感染ウイルスを得ることが可能である (例えば、Logan, J. 及び T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81: 3655-3659 を参照) 。更に、ラウス肉腫ウイルス (RSV) エンハンサー等の転写エンハンサーを用いて、哺乳動物宿主細胞における発現を増大させ得る。SV40またはEBVをベースにしたベクターを用いてタンパク質を高レベルで発現させることもできる。

【0136】

ヒト人工染色体 (HAC) を用いて、プラスミドに含まれ且つプラスミドから発現するものより大きなDNAの断片を輸送することもできる。治療のために約6kb~10MbのHACsを作製し、従来の輸送方法 (リポソーム、ポリカチオンアミノポリマー、またはベシクル) で供給する (Harrington, J. J. ら (1997) *Nat Genet.* 15: 345-355 を参照)。

【0137】

長期にわたり哺乳動物系内で組換えタンパク質を生成するためには、細胞株内のGCRECの安定発現が望ましい。例えば、発現ベクターであって複製発現因子、内在性発現因子、或いはその両者のウイルス起源を含むものと、同一或いは別のベクター上の選択可能マーカー遺伝子とを用いて、GCRECをコードする配列を細胞株に形質転換することが可能である。ベクターの導入後、選択培地に移す前に強化培地で約1~2日間細胞を増殖させることができる。選択可能マーカーの目的は選択培地への抵抗性を与えることであり、選択可能マーカーが存在することにより、導入された配列をうまく発現するような細胞の

成長及び回収が可能となる。安定的に形質転換された細胞の耐性クローンは、その細胞型に適した組織培養技術を用いて増殖可能である。

【0138】

任意の数の選択系を用いて、形質転換細胞株を回収できる。限定するものではないがこのような選択系には、*tk*⁻単細胞のために用いられる単純ヘルペスウイルスチミジンキナーゼ遺伝子と、*apr*⁻細胞のために用いられるアデニンホスホリボシルトランスフェラーゼ遺伝子がある(例えば、Wigler, M. 他(1977) *Cell* 11: 223-232、Lowy, I. 他(1980) *Cell* 22: 817-823. 等を参照)。また、選択の基礎として代謝拮抗物質、抗生物質或いは除草剤への耐性を用いることができる。例えば *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* はクロルスルフロロンに対する耐性を、*pat* はホスフィノトリシンアセチルトランスフェラーゼに対する耐性を各々与える(Wigler, M. 他(1980) *Proc. Natl. Acad. Sci. USA* 77: 3567-3570、Colbere-Garapin, F. 他(1981) *J. Mol. Biol.* 150: 1-14 等を参照)。その他の選択可能な遺伝子、例えば、代謝のための細胞の必要条件を変える *trpB* 及び *hisD* は、文献に記載されている(Hartman, S.C. 及び R.C. Mulligan(1988) *Proc. Natl. Acad. Sci. USA* 85: 8047-8051 等を参照)。可視マーカー、例えばアントシアニン、緑色蛍光タンパク質(GFP; Clontech)、グルクロニダーゼ及びその基質グルクロニド、またはルシフェラーゼ及びその基質ルシフェリン等を用いてもよい。これらのマーカーを用いて、トランスフォーマントを特定するだけでなく、特定のベクター系に起因する一過性或いは安定したタンパク質発現を定量することが可能である(Rhodes, C.A. (1995) *Methods Mol. Biol.* 55: 121-131 等を参照)。

10

20

【0139】

マーカー遺伝子発現の存在/不存在によって目的の遺伝子の存在が示唆されても、その遺伝子の存在及び発現の確認が必要な場合もある。例えば、GCRECをコードする配列がマーカー遺伝子配列内に挿入された場合、GCRECをコードする配列を含む形質転換細胞は、マーカー遺伝子機能の欠落により同定することが可能である。または単一プロモーター制御下で、GCRECをコードする配列とタンデムにマーカー遺伝子を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。

30

【0140】

一般に、GCRECをコードする核酸配列を含み且つGCRECを発現する宿主細胞は、当業者によく知られている種々の方法を用いて同定することが可能である。限定するものではないが当業者によく知られている方法には、DNA-DNA或いはDNA-RNAハイブリダイゼーション、PCR法、核酸或いはタンパク質の検出、定量、或いはその両方を行うための膜系、溶液ベース或いはチップベースの技術を含むタンパク質バイオアッセイまたはイムノアッセイ技術がある。

【0141】

特異的ポリクローナル抗体または特異的モノクローナル抗体を用いてGCRECの発現の検出及び計測を行うための免疫学的方法は、当分野で公知である。このような技術の例としては、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)、蛍光活性化細胞選別(FACS)などが挙げられる。GCREC上の2つの非干渉エピトープに反応するモノクローナル抗体を用いた、2部位モノクローナルベースのイムノアッセイ(two-site, monoclonal-based immunoassay)が好ましいが、競合結合アッセイも用いることもできる。これらのアッセイ及びこれ以外のアッセイは、当分野で公知である(Hampton, R. 他(1990) *Serological Methods, a Laboratory Manual*. APS Press. St Paul, MN, Sect. IV, Coli

40

50

gan, J. E. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY, Pound, J. D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ等を参照)。

【0142】

多岐にわたる標識方法及び結合方法が、当業者に知られており、様々な核酸アッセイおよびアミノ酸アッセイにこれらの方法を用い得る。GCREC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プローブを産出する方法には、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増幅法がある。或いは、GCRECをコードする配列またはその任意の断片を、mRNAプローブを産出するためのベクターにクローニングすることも可能である。このようなベクターは、当分野において知られており、市販もされており、T7、T3またはSP6等の好適なRNAポリメラーゼ及び標識されたヌクレオチドを加えて、in vitroでRNAプローブの合成に用いることができる。このような方法は、例えばAmersham Pharmacia Biotech、Promega (Madison WI)、U.S. Biochemical等から市販されている種々のキットを用いて実行することができる。検出を容易にするために用い得る好適なレポーター分子或いは標識には、基質、補助因子、インヒビター、磁気粒子のほか、放射性核種、酵素、蛍光剤、化学発光剤、発色剤等がある。

10

20

【0143】

GCRECをコードするヌクレオチド配列を用いて形質転換した宿主細胞は、細胞培地からのタンパク質の回収及び発現に適した条件下で培養し得る。形質転換細胞から製造されたタンパク質が分泌されるか細胞内に留まるかは、使用される配列、ベクター、或いはその両者に依存する。当業者であれば理解し得るように、GCRECをコードするポリヌクレオチドを含む発現ベクターを、原核細胞膜または真核細胞膜を透過するGCRECの分泌を誘導するシグナル配列を含むように設計し得る。

【0144】

更に、宿主細胞株の選択は、挿入した配列の発現を調節する能力または発現したタンパク質を所望の形に処理する能力によって行い得る。限定するものではないがこのようなポリペプチドの修飾には、アセチル化、カルボキシル化、グリコシル化、リン酸化、脂質化及びアシル化がある。タンパク質の「プレプロ」または「プロ」形を切断するような翻訳後処理を利用して、タンパク質のターゲティング、折りたたみ及び/または活性を特定することも可能である。翻訳後の活性のための固有の細胞装置及び特徴のある機構を有する種々の宿主細胞(例えばCHO、HeLa、MDCK、MEK293、WI38等)は、American Type Culture Collection (ATCC, Manassas, VA)から入手可能であり、外来タンパク質の正しい修飾及び処理を確実にするように選択し得る。

30

【0145】

本発明の別の実施例では、GCRECをコードする天然の核酸配列、修飾核酸配列または組換え核酸配列を、上記任意の宿主系において融合タンパク質の翻訳をもたらす異種配列に連結反応させることができる。例えば、市販されている抗体を用いて認識可能な異種部分を含むキメラGCRECタンパク質は、GCREC活性阻害剤に対するペプチドライブラリのスクリーニングを促進し得る。また、異種タンパク質部分及び異種ペプチド部分も、市販されている親和性基質を用いて融合タンパク質の精製を促進し得る。限定されるものではないがこのような部分には、グルタチオンSトランスフェラーゼ(GST)、マルトース結合タンパク質(MBP)、チオレドキシン(Trx)、カルモジュリン結合ペプチド(CBP)、6-His、FLAG、c-myc、赤血球凝集素(HA)がある。GSTは固定化グルタチオン上で、MBPはマルトース上で、Trxはフェニルアルシンオキシド上で、CBPはカルモジュリン上で、そして6-Hisは金属キレート樹脂上で、

40

50

同族の融合タンパク質の精製を可能にする。FLAG、c-myc及び赤血球凝集素(HA)は、これらのエピトープ標識を特異的に認識する市販されているモノクローナル抗体及びポリクローナル抗体を用いて、融合タンパク質の免疫親和性精製を可能にする。また、融合タンパク質を遺伝子操作し、GCRECが精製後に異種部分から切断され得るように、GCRECコード配列と異種タンパク質配列の間にタンパク質分解性開裂部位を含めることもできる。融合タンパク質の発現及び精製方法は、前出のAusubel(1995) 10章に記載されている。市販されている種々のキットを用いて融合タンパク質の発現及び精製を促進することもできる。

【0146】

本発明の更に別の実施例では、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、放射能標識したGCRECの合成が*in vitro*で可能である。これらの系は、T7、T3またはSP6プロモーターと機能的に結合したタンパク質コード配列の転写及び翻訳を結合する。翻訳は、例えば³⁵Sメチオニンのような放射能標識したアミノ酸前駆体の存在下で起こる。

【0147】

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。

【0148】

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に擬似のパートナーまたは自然に結合する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、ショウジョウバエ、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。

【0149】

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

【0150】

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*または無細胞再構成系と結合させてアッセイを

10

20

30

40

50

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。

【0151】

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(Martha, 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、可能性のある治療薬や毒性薬剤で検査することができる。

10

20

【0152】

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)。

【0153】

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、可能性のある医薬品を用いて処理し、ヒトの疾患の治療に関する情報を得る。別法では、例えば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)。

30

【0154】

治療

化学的及び構造的類似性、例えば配列及びモチーフとの関連における類似性が、GCRECの領域とGタンパク質結合受容体間に存在する。更に、GCRECの発現は乳房組織、巨核芽球細胞、前立腺腫瘍、後根神経節組織、脳下垂体腺組織に密接に関連する。従ってGCRECは、細胞増殖異常、神経障害、心血管障害、胃腸障害、自己免疫/炎症疾患及び代謝障害並びにウイルス感染症において或る役割を果たすものと考えられる。GCRECの発現また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。

40

【0155】

従って、或る実施例において、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

50

ある。限定するものではないがこのような疾患の例として細胞増殖異常が含まれ、その中には日光性角化症、動脈硬化症、アテローム性動脈硬化症、滑液包炎、硬変、肝炎、混合型結合組織病(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、原発性血小板血症と、腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫、奇形癌を含む癌、具体的には副腎、膀胱、骨、骨髄、脳、乳房、頸部、胆嚢、神経節、消化管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、子宮の癌等が含まれ、神経障害が含まれ、その中には癲癇、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、パーキンソン病その他の錐体外路障害、筋萎縮性側索硬化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、多発性硬化症その他の脱髄疾患、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、クロイツフェルトヤコブ病及びガスマンストラウラーシャインカー症候群を含むプリオン病と、致死性家族性不眠症、神経系の栄養病及び代謝病、神経線維腫症、結節硬化症、小脳網膜性血管芽腫症(cerebelloretinal hemangioblastomatosis)、脳3叉神経血管症候群、精神薄弱その他の中枢神経系発達障害、脳性麻痺、神経骨格異常、自律神経系障害、脳神経障害、脊髄病、筋ジストロフィーその他の神経筋疾患、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、気分障害、不安障害及び精神分裂病を含む精神障害と、季節型感情障害(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、トゥレット病、進行性核上麻痺、皮質基底核変性症(corticobasal degeneration)、家族性前頭側頭骨痴呆が含まれ、また心血管障害も含まれ、その中には動静脈瘤、アテローム性動脈硬化症、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、血栓崩壊の合併症、バルーン血管形成術(balloon angioplasty)、血管置換術、大動脈冠動脈バイパス術移植手術(coronary artery bypass graft surgery)などの血管疾患と、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、全身性紅斑性狼瘡の心内膜炎、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、心臓移植の合併症などの心疾患が含まれ、また胃腸障害も含まれ、その中には嚥下障害、消化性食道炎、食道痙攣、食道狭窄、食道癌、消化不良、消化障害、胃炎、胃癌、食欲不振、悪心、嘔吐、胃不全麻痺、洞または幽門の浮腫、腹部アングナ、胸焼け、胃腸炎、イレウス、腸管感染、消化性潰瘍、胆石症、胆嚢炎、胆汁うっ滞、膵臓炎、膵臓癌、胆道疾患、肝炎、高ビリルビン血症、硬変症、肝臓の受動性うっ血、ヘパトーム、感染性大腸炎、潰瘍性大腸炎、潰瘍性直腸炎、クローン病、ホイップル病、マロリーヴァイス症候群、結腸癌、結腸閉塞、過敏性腸症候群、短小腸症候群、下痢、便秘、胃腸出血、後天性免疫不全症候群(AIDS)腸症、黄疸、肝性脳症、肝腎症候群、肝炎、血色素症、ウィルソン病、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、癌腫を含む肝癌が含まれ、また自己免疫/炎症疾患も含まれ、その中には後天性免疫不全症候群(AIDS)、アジソン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫多発性内分泌腺障害症(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

10

20

30

40

50

大腸症候群、多発性硬化症、重症筋無力症、心筋または心膜の炎症、骨関節炎、骨粗しょう症、膵炎、多発性筋炎、乾癬、ライター症候群、リウマチ様関節炎、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性紅斑性狼瘡、全身性硬化症、血小板減少症、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌の合併症、血液透析、体外循環、ウイルス性感染症、細菌性感染症、真菌性感染症、寄生虫感染症、原虫感染症、蠕虫の感染症及び外傷が含まれ、また代謝障害も含まれ、その中には糖尿病、肥満症及び骨粗鬆症が含まれ、またアデノウイルス、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、パラミクソウイルス、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、ラプトウイルス及びトンガウイルス (t o n g a v i r u s) として分類されるウイルス性病原体による感染症も含まれる。

10

【0156】

別の実施例では、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して、限定するものではないが上記した疾患を含む G C R E C の発現または活性の低下に関連した疾患を治療または予防することも可能である。

【0157】

更に別の実施例では、実質的に精製された G C R E C を含む成分を好適な医薬用担体と共に患者に投与して、限定するものではないが上記した疾患を含む G C R E C の発現または活性の低下に関連した疾患を治療または予防することも可能である。

20

【0158】

更に別の実施例では、G C R E C の活性を調節するアゴニストを患者に投与して、限定するものではないが上記した疾患を含む G C R E C の発現または活性の低下に関連した疾患を治療または予防することも可能である。

【0159】

更に別の実施例では、患者に G C R E C のアンタゴニストを投与して、G C R E C の発現または活性の増大に関連した疾患を治療または予防することが可能である。限定するものではないがこのような疾患の例には、上記した細胞増殖異常、神経障害、心血管障害、胃腸障害、自己免疫/炎症疾患及び代謝障害並びにウイルス性感染症がある。一実施態様においては、アンタゴニストとして直接的に、或いは G C R E C を発現する細胞または組織に薬剤を輸送するターゲティングまたは輸送機構として間接的に G C R E C と特異結合する抗体を用いることができる。

30

【0160】

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の相補体を発現するベクターを患者に投与して、限定するものではないが上記した疾患を含む G C R E C の発現または活性の増大に関連した疾患を治療または予防することも可能である。

【0161】

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

40

【0162】

G C R E C のアンタゴニストは、本技術分野で一般的に知られている方法を用いて製造し得る。具体的には、精製された G C R E C を用いて抗体を作るか、治療薬のライブラリをスクリーニングして、G C R E C と特異結合するものを同定することが可能である。G C R E C の抗体も、本技術分野で一般的に知られている方法を用いて製造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抗体には、ポリクローナル抗体、モノクローナル抗体、キメラ抗体、一本鎖抗体、F a b 断片及び F a b 発現ライブラリによって作

50

られた断片が含まれ得る。中和抗体（即ち二量体の形成を阻害する抗体）は通常、治療用に好適である。

【0163】

抗体の産生のためには、ヤギ、ウサギ、ラット、マウス、ヒト、その他を含む種々の宿主は、GCREC、またはGCRECの任意の断片またはオリゴペプチドであって免疫抗原性の特性を有するものを注入することによって免疫化することができる。宿主の種に応じて、種々のアジュバントを用いて免疫応答を高めることもできる。限定するものではないがこのようなアジュバントには、フロイントアジュバントと、水酸化アルミニウム等のミネラルゲルアジュバントと、リゾレシチン、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、ジニトロフェノール等の界面活性剤とがある。ヒトに用いられるアジュバントの中では、BCG（カルメット グラン杆菌）及びコリネバクテリウム パルヴムが特に好ましい。

10

【0164】

GCRECに対する抗体を誘導するために用いるオリゴペプチド、ペプチドまたは断片は、少なくとも約5のアミノ酸からなり、一般的には少なくとも約10のアミノ酸からなるアミノ酸配列を有するものが好ましい。これらのオリゴペプチド、ペプチドまたは断片は、天然のタンパク質のアミノ酸配列の一部と同一であることが望ましい。GCRECアミノ酸の短い伸長部は、別のタンパク質、例えばキーホールリンペットヘモシニアンの配列と融合し、キメラ分子に対する抗体が産生され得る。

【0165】

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

20

【0166】

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、Morrisson, S. L. 他. (1984) Proc. Natl. Acad. Sci. 81: 6851-6855、Neuberger, M. S. 他. (1984) Nature 312: 604-608; Takeda, S. ら. (1985) Nature 314: 452, 454等を参照）。或いは、一本鎖抗体を産生するために説明された技術を適用し、本技術分野で知られている方法を用いて、GCREC特異性一本鎖抗体を産生し得る。関連特異性を有するガイディオタイプ組成が異なるような抗体を、ランダムな組合せの免疫グロブリンライブラリからチェーンシャッフリングによって産生することもできる（Burton D. R. (1991) Proc. Natl. Acad. Sci. USA 88: 10134-10137等を参照）。

30

40

【0167】

抗体の産生は、リンパ球集団における *in vivo* 産生の誘導によって、或いは免疫グロブリンライブラリのスクリーニングまたは文献に開示されているような高特異結合試薬のパネルのスクリーニングによっても行い得る（Orlandi, R. ら (1989) Proc. Natl. Acad. Sci. USA 86: 3833-3837、Winter, G. ら (1991) Nature 349: 293-299等を参照）。

【0168】

GCRECのための特異結合部位を有する抗体を産生することもできる。例えば、限定す

50

るものではないが、このような断片には、抗体分子のペプシン消化によって作製される $F(ab')_2$ 断片と、 $F(ab')_2$ 断片のジスルフィド架橋を還元することによって作製される Fab 断片とがある。或いは、Fab 発現ライブラリを作製することによって、モノクローナル Fab 断片を所望の特異性と迅速且つ容易に同定することが可能となる (Huse, W. D. ら (1989) *Science* 246: 1275-1281 等を参照)。

【0169】

種々のイムノアッセイを用いてスクリーニングし、所望の特異性を有する抗体を同定することができる。確立された特異性を有するポリクローナル抗体またはモノクローナル抗体の何れかを用いる免疫放射線アッセイまたは競合結合アッセイに対する数々のプロトコルは、本技術分野において公知である。このようなイムノアッセイは通常、GCREC とその特異性抗体間の複合体形成の計測に關与している。2つの非干渉性 GCREC エピトープに反応するモノクローナル抗体を用いるような、2部位モノクローナルベースのイムノアッセイが一般に利用されるが、競合結合アッセイを利用してよい (前出の Pound の文献)。

10

【0170】

ラジオイムノアッセイ技術と共に様々な方法、例えばスキッチャード分析を用いて、GCREC に対する抗体の親和性を評価する。親和性は結合定数 K_a で表す。 K_a は、平衡状態において 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た値であると定義する。ポリクローナル抗体は多様な GCREC エピトープに対する親和性が不均一であり、ポリクローナル抗体試薬のために決定した K_a は、GCREC 抗体の平均親和性または結合活性を表す。モノクローナル抗体は特定の GCREC エピトープに対して単一特異的であり、モノクローナル抗体試薬のために決定した K_a は、親和性の真の測定値を表す。 K_a 値が約 $10^9 \sim 10^{12}$ L/mol の範囲にあるような高親和性抗体試薬は、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。 K_a 値が $10^6 \sim 10^7$ L/mol の低親和性抗体試薬は、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 及び Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY)。

20

30

【0171】

ポリクローナル抗体試薬の抗体価及び結合活性を更に評価して、後に使う或る適用例に対するこのような試薬の品質及び適性を決定することができる。例えば、少なくとも $1 \sim 2$ mg/ml、好ましくは $5 \sim 10$ mg/ml の特異抗体を含むポリクローナル抗体試薬は、GCREC 抗体複合体を沈殿させる必要がある処理において通常用いられる。抗体の特異性、抗体価、結合活性、様々な適用例における抗体の品質や使用に対する指針については、一般に入手可能である。(前出の Catty の文献、同 Coligan らの文献等を参照)。

40

【0172】

本発明の別の実施例では、GCREC をコードするポリヌクレオチド、GCREC の任意の断片または相補配列を治療目的で使用することができる。ある実施態様では、GCREC をコードする遺伝子のコーディング領域や調節領域に相補的な配列やアンチセンス分子 (DNA 及び RNA、PNA、修飾オリゴヌクレオチド) を設計して遺伝子発現を変更することができる。このような技術は当分野では周知であり、アンチセンスオリゴヌクレオチドまたは大きな断片が、GCREC をコードする配列の制御領域から、またはコード領域に沿ったさまざまな位置から設計可能である。(Agrawal, S 編集. (1996) Antisense Therapeutics, Humana Press In

50

c. , T o t a w a N J を参照)。

【0173】

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輸送することが可能である (Slater, J. E. ら (1998) *J. Allergy Clin. Immunol.* 102(3): 469 - 475、Scanlon, K. J. ら (1995) 9(13): 1288 - 1296. 等を参照)。アンチセンス配列はまた、例えばレトロウイルスやアデノ関連ウイルスベクター等のウイルスベクターを用いて細胞内に導入することもできる (Miller, A. D. (1990) *Blood* 76: 271、前出のAusubel、Uckert, W及びW. Walther (1994) *Pharmacol. Ther.* 63(3): 323 - 347等を参照)。その他の遺伝子送達機構には、リポソーム系、人工的なウイルスエンベロープ及び当分野で公知のその他の系が含まれる (Rossi, J. J. (1995) *Br. Med. Bull.* 51(1): 217 - 225; Boado, R. J. ら (1998) *J. Pharm. Sci.* 87(11): 1308 - 1315、Morris, M. C. ら (1997) *Nucleic Acids Res.* 25(14): 2730 - 2736. 等を参照)。

10

【0174】

本発明の別の実施例では、GCRECをコードするポリヌクレオチドを、体細胞若しくは生殖細胞遺伝子治療に用いることが可能である。遺伝子治療を行うことにより、(i) 遺伝子欠損症 (例えばX染色体鎖遺伝 (Cavazzana-Calvo, M. ら (2000) *Science* 288: 669 - 672) により特徴付けられる重度の複合型免疫欠損 (SCID) - X1の場合)、先天性アデノシンデアミナーゼ (ADA) 欠損症に関連する重度の複合型免疫欠損 (Blaese, R. M. ら (1995) *Science* 270: 475 - 480、Bordignon, C. ら (1995) *Science* 270: 470 - 475)、嚢胞性繊維症 (Zabner, J. ら (1993) *Cell* 75: 207 - 216; Crystal, R. G. ら (1995) *Hum. Gene Therapy* 6: 643 - 666、Crystal, R. G. ら (1995) *Hum. Gene Therapy* 6: 667 - 703)、サラセミア (thalassaemia)、家族性高コレステロール血症、第VIII因子若しくは第IX因子欠損に起因する血友病 (Crystal, R. G. (1995) *Science* 270: 404 - 410、Verma, I. M. 及びSomia, N. (1997) *Nature* 389: 239 - 242) を治療し、(ii) 条件的致死性遺伝子産物を発現させ (例えば制御不能な細胞増殖に起因する癌の場合)、(iii) 細胞内の寄生虫 (例えばヒト免疫不全ウイルス (HIV) (Baltimore, D. (1988) *Nature* 335: 395 - 396、Poeschl, E. ら (1996) *Proc. Natl. Acad. Sci. USA.* 93: 11395 - 11399)、B型若しくはC型肝炎ウイルス (HBV、HCV)、口腔カンジダ症菌及び *Paracoccidioides brasiliensis* 等の真菌寄生虫、並びに熱帯熱マラリア原虫及びクルーズトリパノソーマ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ることができる。GCRECの発現若しくは調節に必要な遺伝子の欠損が疾患を発生させる場合、形質導入した細胞の好適な集団からGCRECを発現することにより、遺伝子欠損に起因する症状の発現を緩和し得る。

20

30

40

【0175】

本発明の更なる実施例では、GCRECをコードする哺乳動物発現ベクターを作製し、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、GCRECの欠損による疾患や異常症を治療する。*in vivo* 或いは *ex vivo* の細胞に用いる機械的導入技術には、(i) 個々の細胞内への直接的なDNA微量注射法、(ii) 遺伝子銃、(iii) リポソームを介した形質移入、(iv) 受容体を介した遺伝

50

子導入、及び(v)DNAトランスポソンの使用(Morgan, R. A. およびW. F. Anderson (1993) *Annu. Rev. Biochem.* 62: 191-217、Ivics, Z. (1997) *Cell* 91: 501-510; Boulay, J-L. およびH. Recipon (1998) *Curr. Opin. Biotechnol.* 9: 445-450)がある。

【0176】

限定するものではないがGCRECの発現に影響を及ぼし得る発現ベクターには、PCDNA3.1、EPITAG、PRCCMV2、PREP、PVAXベクター(Invitrogen, Carlsbad CA)、PCMV-SCRIPT、PCMV-TAG、PEGSH/PERV(Stratagene, La Jolla CA)及びPTE-T-OFF、PTET-ON、PTRE2、PTRE2-LUC、PTK-HYG(Clontech, Palo Alto CA)がある。GCRECを発現させるために、(i)恒常的に活性なプロモーター(例えば、サイトメガロウイルス(CMV)、ラウス肉腫ウイルス(RSV)、SV40ウイルス、チミジンキナーゼ(TK)、若しくは - アクチン遺伝子等)、(ii)誘導性プロモーター(例えば、市販されているT-REXプラスミド(Invitrogen)に含まれている、テトラサイクリン調節性プロモーター(Gossen, M. 及びH. Bujard (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89: 5547-5551; Gossen, M. 他 (1995) *Science* 268: 1766-1769; Rossi, F.M.V. 及びH.M. Blau (1998) *Curr. Opin. Biotechnol.* 9: 451-456)、エクジソン誘導性プロモーター(市販されているプラスミドPVGRXR及びPINDに含まれている: Invitrogen)、FK506/ラパマイシン誘導性プロモーター、またはRU486/ミフェプリストーン誘導性プロモーター(Rossi, F.M.V. 及びH.M. Blau, 前出)、または(iii)正常な個体に由来する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。

【0177】

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

【0178】

本発明の別の実施例では、GCRECの発現に関連する遺伝子欠損によって起こる疾患や異常症は、(i)レトロウイルス末端反復配列(LTR)プロモーターまたは独立プロモーターの制御下でGCRECをコードするポリヌクレオチドと、(ii)好適なRNAパッケージングシグナルと、(iii)追加レトロウイルス・シス作用性RNA配列及び効率的なベクターの増殖に必要なコード配列を伴うRev応答性エレメント(RRE)とからなるレトロウイルスベクターを作製して治療することができる。レトロウイルスベクター(例えばPFB及びPFBNEO)はStratagene社から市販されており、刊行データ(Riviere, I. ら (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。ベクターは、好適なベクター産生細胞系(VPCL)において増殖され、VPCLは、標的細胞上の受容体に対する向性を有するエンベロープ遺伝子またはVSVg等の乱交雑エンベロープタンパク質を発現する(Armentano, D. ら (1987) *J. Virol.* 61: 1647-1650、Bender, M.A. ら (1987) *J. Virol.* 61: 1639-164

6、Adam, M. A. 及び A. D. Miller (1988) J. Virol. 62:3802-3806、Dull, T. ら (1998) J. Virol. 72:8463-8471、Zufferey, R. ら (1998) J. Virol. 72:9873-9880)。RIGGに付与された米国特許第5,910,434号(「Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant」)において、レトロウイルスパッケージング細胞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、細胞集団(例えばCD4⁺T細胞)の形質導入、及び形質導入した細胞の患者への戻しは、遺伝子治療の分野では当業者に公知の方法であり、多数の文献に記載されている(Ranga, U. ら (1997) J. Virol. 71:7020-7029、Bauer, G. ら (1997) Blood 89:2259-2267、Bonyhadi, M. L. (1997) J. Virol. 71:4707-4716、Ranga, U. ら (1998) Proc. Natl. Acad. Sci. U.S.A. 95:1201-1206、Su, L. (1997) Blood 89:2283-2290)。

【0179】

別法では、アデノウイルス系遺伝子治療の輸送系を用いて、GCRECの発現に関連する1つ若しくは複数の遺伝子異常を有するような細胞にGCRECをコードするポリヌクレオチドを輸送する。アデノウイルス系ベクターの作製及びパッケージングについては、当業者に公知である。複製欠損型アデノウイルスベクターは、免疫調節タンパク質をコードする遺伝子を膵臓の無損傷の膵島内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. ら (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は、Armentanoに付与された米国特許第5,707,618号(「Adenovirus vectors for gene therapy」)に記載されており、引用することを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、Antinozzi, P. A. ら (1999) Annu. Rev. Nutr. 19:511-544、Verma, I. M. 及び N. Somia (1997) Nature 18:389:239-242等も参照されたい。両文献は、引用することをもって本明細書の一部とする。

【0180】

別法では、ヘルペス系遺伝子治療の輸送系を用いて、GCRECの発現に関連する1つ若しくは複数の遺伝子異常を有する標的細胞にGCRECをコードするポリヌクレオチドを輸送する。HSVが向性を有するような中枢神経系の細胞にGCRECを導入する際には、単純ヘルペスウイルス(HSV)系のベクターの使用は特に役立つ。ヘルペス系ベクターの作製及びパッケージングは、当業者に公知である。複製適格性単純ヘルペスウイルス(HSV)I型系のベクターは、レポーター遺伝子を霊長類の眼に送達するために用いられてきた(Liu, X. ら (1999) Exp. Eye Res. 169:385-395)。HSV-1ウイルスベクターの作製についても、DeLucaに付与された米国特許第5,804,413号(「Herpes simplex virus swains for gene transfer」)が開示されており、該特許の引用をもって本明細書の一部とする。米国特許第5,804,413号には、ヒト遺伝子治療を含む目的のために好適なプロモーターの制御下において細胞に導入される少なくとも1つの外在性遺伝子を有するゲノムを含む組換えHSV d92についての記載がある。上記特許はまた、ICP4、ICP27及びICP22のために除去される組換えHSV系統の作製及び使用について開示している。HSVベクターについては、Goins, W. F. ら (1999) J. Virol. 73:519-532 及び Xu, H. ら (1994) Dev. Biol. 163:152-161も参照されたい。両文献は、引用をもって本明細書の一部とする。クローン化ヘルペスウイルス配列の操作、巨大ヘル

ペスウイルスのゲノムの異なった部分を含む多数のプラスミドを形質移入した後の組換えウイルスの産生、ヘルペスウイルスの成長及び増殖、並びにヘルペスウイルスの細胞への感染は、当業者に公知の技術である。

【0181】

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輸送する。プロトタイプのウイルスであるセムリキ森林熱ウイルス（Semliki Forest Virus, SFV）の生物学的研究が広範に行われており、遺伝子導入ベクターがSFVゲノムに基づいていることが分かった（Garoff, H. および K.-J. Li (1998) *Cun. Opin. Biotech.* 9: 464 - 469）。ウイルスRNAの複製中に、通常はウイルスのキャプシッドタンパク質をコードするサブゲノムRNAが作り出される。このサブゲノムRNAは、完全長のゲノムRNAより高いレベルに複製されるため、酵素活性（例えばプロテアーゼ及びポリメラーゼ）を有するウイルスタンパク質に比べてキャプシッドタンパク質が過剰産生される。同様に、GCRECに対するコード配列をカプシッドコード領域のウイルスゲノムに導入することにより、ベクター導入細胞において多数のGCRECコードRNAが産生され、高レベルのGCRECが合成される。通常はウイルスの感染が数日以内での細胞溶解に関係する一方で、シンドビスウイルス（SIN）の変異体を有するハムスター正常腎臓細胞（BHK-21）の持続的な感染を確立する能力は、ウイルスの溶解複製を遺伝子治療に適用できるように好適に変更可能であることを示唆している（Dryga, S.A. ら (1997) *Virology* 228: 74 - 83）。ウイルスの宿主の範囲が広いことにより、様々な細胞タイプへのGCRECの導入が可能になる。或る集団におけるサブセットの細胞の特定形質導入は、形質導入前に細胞の選別を必要とし得る。ウイルスの感染性cDNAクローンの処置方法、ウイルスのcDNA及びRNAの形質移入方法及びウイルスの感染方法は、当業者に公知である。

10

20

【0182】

転写開始部位由来のオリゴヌクレオチドを用いて遺伝子発現を阻害することも可能である。転写開始部位とは例えば開始部位から数えて約 - 10 と約 + 10 の間である。同様に、三重らせん塩基対の形成方法を用いて阻害が可能となる。三重らせん塩基対形成は、ポリメラーゼ、転写因子または調節分子の結合のために十分に開くような二重らせんの能力を阻害するので、三重らせん塩基対形成は有用である。三重らせんDNAを用いる最近の治療の進歩については文献に記載がある（Gee, J.E. ら (1994) *in: Huber, B.E. 及び B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY, 163 - 177* ページ等を参照）。相補配列またはアンチセンス分子もまた、転写物がリボソームに結合するのを阻止することによってmRNAの翻訳を阻止するように設計することができる。

30

【0183】

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

40

【0184】

任意の潜在的RNA標的内の特異的リボザイム切断部位は、GUA、GUU、GUC配列を含めたりボザイム切断部位に対する標的分子をスキャンすることによって先ず同定される。一度同定すると、オリゴヌクレオチドを機能不全にするような2次構造の特徴に対して切断部位を含む標的遺伝子の領域に対応する15～20リボヌクレオチドの短いRNA配列を、評価することが可能になる。候補標的の適合性の評価も、リボヌクレアーゼ保護アッセイを用いて相補的オリゴヌクレオチドとのハイブリダイゼーションの実施容易性を

50

テストすることによって行うことができる。

【0185】

本発明の相補リボ核酸分子及びリボザイムは、核酸分子合成のために当分野でよく知られている任意の方法を用いて作製し得る。任意の方法には、固相フォスフォアミダイト化合物等のオリゴヌクレオチドを化学的に合成する方法がある。或いは、GCRECをコードするDNA配列の*in vitro*及び*in vivo*転写によってRNA分子を産出し得る。このようなDNA配列は、T7やSP6等の好適なRNAポリメラーゼプロモーターを用いて多様なベクター内に取り込むことが可能である。或いは、相補的RNAを構成的或いは誘導的に合成するようなこれらcDNA産物を、細胞株、細胞または組織内に導入することができる。

10

【0186】

細胞内の安定性を高め、半減期を長くするためにRNA分子を修飾することができる。限定するものではないが可能な修飾には、分子の5'末端、3'末端、あるいはその両方においてフランキング配列を追加したり、分子の主鎖内においてホスホジエステラーゼ結合ではなくホスホチオネートまたは2'-O-メチルを使用したりすることが含まれる。この概念は、PNAの産出に固有のものであり、これら全ての分子に拡大することができる。それには、内在性エンドヌクレアーゼによって容易には認識されないアデニン、シチジン、グアニン、チミン、及びウリジンにアセチル-、メチル-、チオ-及び同様の修飾をしたものの他、非従来型塩基、例えばイノシン(queosine)、ワイプトシン(wybutosine)等を加えることができる。

20

【0187】

本発明の追加実施例は、GCRECをコードするポリヌクレオチドの変異発現に有効な化合物をスクリーニングする方法を含む。限定するものではないが特異ポリヌクレオチドの変異発現に有効な化合物には、オリゴヌクレオチド、アンチセンスオリゴヌクレオチド、三重らせん形成オリゴヌクレオチド、転写因子その他のポリペプチド転写制御因子、及び特異ポリヌクレオチド配列と相互作用し得る非高分子化学的実体がある。有効な化合物は、ポリヌクレオチド発現のインヒビターまたはプロモーターのいずれかとして作用することによりポリヌクレオチド発現を変異し得る。従って、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。

30

【0188】

特異ポリヌクレオチドの変異発現における有効性に対して、少なくとも1個から複数個の試験化合物をスクリーニングし得る。試験化合物は、当分野で通常知られている任意の方法により得られる。このような方法には、ポリヌクレオチドの発現を変異させる場合と、既存の、市販のまたは専売の、天然または非天然の化合物ライブラリから選択する場合と、標的ポリヌクレオチドの化学的及び/または構造的特性に基づく化合物を合理的にデザインする場合と、組合せ的にまたは無作為に生成した化合物のライブラリから選択する場合に有効であることが知られているような化合物の化学修飾がある。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され、このように得られる。サンプルには例えば、無傷細胞、透過化処理した細胞、無細胞再構成系または再構成生化学系があり得る。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、当分野で通常知られている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、特異ヌクレオチドの発現を検出する。ハイブリダイゼーション量を定量し、それによって1つ以上の試験化合物に曝露される及び曝露されないポリヌクレオチドの発現の比較に対する基礎を形成し得る。試験化合物に曝露されるポリヌクレオチドの発現における変化の検出は、ポリヌクレオチドの発現を変異する際に試験化合物が有効であることを示している。特異ポリヌクレオチドの変異発現に有効な化合

40

50

物に対して、例えば Schizosaccharomyces pombe 遺伝子発現系 (Atkins, D. ら (1999) 米国特許第 5,932,435 号、Arndt, G.M. ら (2000) Nucleic Acids Res. 28:E15) または HeLa 細胞等のヒト細胞系 (Clarke, M.L. ら (2000) Biochem. Biophys. Res. Commun. 268:8-13) を用いてスクリーニングを実行する。本発明の特定の実施例は、特異的ポリヌクレオチド配列に対するアンチセンス活性のためのオリゴヌクレオチド (デオキシリボヌクレオチド、リボヌクレオチド、ペプチド核酸、修飾オリゴヌクレオチド) の組み合わせライブラリをスクリーニングすることに関与している (Bruice, T.W. ら (1997) 米国特許第 5,686,242 号、Bruice, T.W. ら (2000) 米国特許第 6,022,691 号)。

【0189】

ベクターを細胞または組織に導入する多数の方法が利用可能であり、in vivo、in vitro 及び ex vivo の使用に対して同程度に適している。ex vivo 治療の場合、ベクターを患者から採取した幹細胞内に導入し、クローニング増殖して同一患者に自家移植で戻すことができる。トランスフェクション、リボソーム注入またはポリカチオンアミノポリマーによる輸送は、当分野でよく知られている方法を用いて実行することができる。(Goldman, C.K. ら (1997) Nat. Biotechnol. 15:462-466. 等を参照)。

【0190】

上記の治療方法はいずれも、例えば、ヒト、イヌ、ネコ、ウシ、ウマ、ウサギ、サル等の哺乳動物を含めて治療が必要な全ての対象に適用できる。

【0191】

本発明の追加実施例は、通常薬剤として許容できる賦形剤で処方される活性成分を有する成分の投与に関連する。賦形剤には例えば、糖、でんぷん、セルロース、ゴム及びタンパク質がある。様々な処方通常知られており、詳細は Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA) の最新版に記載されている。このような成分は、GCREC、GCREC に対する抗体、擬態、アゴニスト、アンタゴニスト、または GCREC インヒビターから構成し得る。

【0192】

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

【0193】

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

【0194】

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

【0195】

特殊形状の成分は、GCREC またはその断片を含む高分子を直接細胞内輸送するために調製される。例えば、細胞不透過性高分子を含むリボソーム製剤は、細胞融合及び高分子の細胞内送達を促進し得る。或いは、GCREC またはその断片を HIV Tat-1 タ

10

20

30

40

50

ンパク質から短い陽イオンN末端部に結合することもできる。このようにして生成された融合タンパク質は、マウスモデル系の脳を含む全ての組織の細胞に形質導入することがわかっている (Schwarze, S. R. ら (1999) Science 285: 1569 - 1572)。

【0196】

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

10

【0197】

治療上の有効投与量は、症状や容態を回復させるような活性成分量を参考にする。そのような活性成分の例としては、GCRECまたはその断片、GCRECの抗体、GCRECのアゴニスト、アンタゴニストまたはインヒビターがある。治療有効度及び毒性は、細胞培養または動物実験における標準的な薬剤手法によって、例えばED₅₀ (集団の50%の医薬的有効量) またはLD₅₀ (集団の50%の致死量) 統計を測定するなどして決定することができる。毒性効果の薬用効果に対する投与量の比は、治療指数であり、LD₅₀ / ED₅₀ 比として表すことができる。高い治療指数を示すような成分が望ましい。細胞培養アッセイ及び動物実験から得られたデータは、ヒトに用いるための投与量の範囲を調剤するのに用いられる。このような組成物が含まれる投与量は、毒性を殆ど或いは全く含まず、ED₅₀ を含むような血中濃度の範囲にあることが好ましい。用いられる投与形態、患者の感受性及び投与の経路によって、投与量はこの範囲内で様々に変わる。

20

【0198】

正確な投与量は、治療が必要な被験者に関する要素を考慮して、現場の医者が決定することになる。効果的なレベルの活性成分を与え、或いは所望の効果を維持するべく、投与量及び投与を調節する。被験者に関する要素としては、疾患の重症度、患者の通常健康状態、患者の年齢、体重及び性別、投与の時間及び頻度、薬剤の配合、反応感受性及び治療に対する応答等を考慮する。作用期間が長い成分は、特定の製剤の半減期及びクリアランス率によって3~4日毎に1度、1週間に1度、或いは2週間に1度の間隔で投与し得る。

30

【0199】

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

【0200】

(診断)

別の実施例では、GCRECの発現によって特徴付けられる疾患の診断のために、或いはGCRECやGCRECのアゴニスト、アンタゴニストまたは阻害剤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おいて、GCRECを特異的に結合する抗体が用いられることがある。診断目的に有用な抗体は、上記の治療の箇所で記載した方法と同じ方法で調合される。GCRECの診断アッセイには、抗体及び標識を利用してヒトの体液において或いは細胞や組織のエキスにおいてGCRECを検出する方法が含まれる。この抗体は修飾されたものもされていないものも可能であり、レポーター分子との共有結合または非共有結合で標識化できる。レポーター分子としては広くさまざまな種類が本分野で知られており、また使用可能であるが、そのうちのいくつかは上記で説明されている。

40

【0201】

GCRECを測定するための様々なプロトコル、例えばELISA、RIA、FACS等

50

が本技術分野において知られており、G C R E C発現の修正レベル或いは異常レベルを診断する基準を提供する。複合体の形成に適した条件下でヒト対象等の正常な哺乳動物対象から採取した体液または細胞抽出とG C R E Cに対する抗体とを結合させることにより、G C R E C発現の正常値または標準値が決定される。標準複合体形成量は、種々の方法、例えば測光法で定量できる。対象内で発現したG C R E Cの量、対照、生検組織からの病変サンプルを標準値と比較する。標準値と被験者との偏差が疾患を診断するパラメータとなる。

【0202】

別の実施例によれば、G C R E Cをコードするポリヌクレオチドを診断目的に用いることもできる。用いることができるポリヌクレオチドには、オリゴヌクレオチド配列、相補的RNA及びDNA分子、そしてPNAが含まれる。ポリヌクレオチドは、検体におけるG C R E Cの発現が疾患と相関し得るような該検体における遺伝子発現の検出及び定量に用いることができる。診断アッセイは、G C R E Cの不在、存在及び過剰発現を測定するために、そして治療インターベンション中にG C R E Cレベルの調製をモニターするために用いることができる。

10

【0203】

ある実施形態では、G C R E Cをコードする核酸配列を同定するために、G C R E Cまたは近縁の分子をコードする、ゲノム配列を含むポリヌクレオチド配列を検出可能なPCRプローブとのハイブリダイゼーションを用いることができる。プローブが同定するのはG C R E C、突然変異体または関連配列をコードするような、天然に存在する配列のみであるか否かは、プローブの特異性及びハイブリダイゼーション或いは増幅のストリンジエンシーによって決定されることになる。ここで、プローブの特異性とは、プローブが高特異領域（例えば5'調節領域）からなるのか、低特異領域（例えば保存されたモチーフ）からなるのかということである。

20

【0204】

プローブはまた、関連する配列の検出にも用いることができ、その配列はG C R E Cをコードする任意の配列と少なくとも50%の相同性を有し得る。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: 18-34の配列、或いはG C R E C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。

30

【0205】

G C R E CをコードするDNAに対する特異的ハイブリダイゼーションプローブを作製する方法には、G C R E CまたはG C R E C誘導体をコードするポリヌクレオチド配列を、mRNAプローブを作製するためのベクターにクローニングする方法が含まれる。mRNAプローブ作製のためのベクターは、当業者に知られており、市販されており、好適な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、i n v i t r oでRNAプローブを合成するために用いられ得る。ハイブリダイゼーションプローブは、種々のレポーターの集団によって標識され得る。レポーター集団の例としては、³²Pまたは³⁵S等の放射性核種、或いはアビジン/ビオチン結合系を介してプローブに結合されたアルカリホスファターゼ等の酵素標識などが挙げられる。

40

【0206】

G C R E Cをコードするポリヌクレオチド配列は、G C R E Cの発現に関係する疾患の診断の為に用い得る。限定するものではないがこのような疾患の例として細胞増殖異常が含まれ、その中には日光性角化症、動脈硬化症、アテローム性動脈硬化症、滑液包炎、硬変、肝炎、混合型結合組織病(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、原発性血小板血症と、腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫、奇形癌を含む癌、具体的には副腎、膀胱、骨、骨髄、脳、乳房、頸部、胆嚢、神経節、消化管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、子宮の癌等が含まれ、神経障害が含まれ、その中には癲癇、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

50

トン病、痴呆、パーキンソン病その他の錐体外路障害、筋萎縮性側索硬化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、多発性硬化症その他の脱髄疾患、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、クロイツフェルト ヤコブ病及びガストマン ストラウスラー シャインカー症候群を含むプリオン病と、致死性家族性不眠症、神経系の栄養病及び代謝病、神経線維腫症、結節硬化症、小脳網膜性血管芽腫症 (cerebelloretinal hemangioblastomatosis)、脳3叉神経血管症候群、精神薄弱その他の中枢神経系発達障害、脳性麻痺、神経骨格異常、自律神経系障害、脳神経障害、脊髄病、筋ジストロフィーその他の神経筋疾患、末梢神経疾患、皮膚筋炎及び多発性筋炎と、遺伝性、代謝性、10
 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、気分障害、不安障害及び精神分裂病を含む精神障害と、季節型感情障害 (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、トゥレット病、進行性核上麻痺、皮質基底核変性症 (corticobasal degeneration)、家族性前頭側頭骨痴呆が含まれ、また心血管障害も含まれ、その中には動静脈瘤、アテローム性動脈硬化症、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、血栓崩壊の合併症、バルーン血管形成術 (balloon angioplasty)、血管置換術、大動脈冠動脈バイパス術移植手術 (coronary artery bypass graft surgery) などの血管疾患と、うっ血性心不全、虚血性心疾患 20
 、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、全身性紅斑性狼瘡の心内膜炎、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、心臓移植の合併症などの心疾患が含まれ、また胃腸障害も含まれ、その中には嚥下障害、消化性食道炎、食道痙攣、食道狭窄、食道癌、消化不良、消化障害、胃炎、胃癌、食欲不振、悪心、嘔吐、胃不全麻痺、洞または幽門の浮腫、腹部アングナ、胸焼け、胃腸炎、イレウス、腸管感染、消化性潰瘍、胆石症、胆嚢炎、胆汁うっ滞、膵臓炎、膵臓癌、胆道疾患、肝炎、高ビリルビン血症、硬変症、肝臓の受動性うっ血、ヘパトーム、感染性大腸炎、潰瘍性大腸炎、潰瘍性直腸炎、クローン病、ホ 30
 イッブル病、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、後天性免疫不全症候群 (AIDS) 腸症、黄疸、肝性脳症、肝腎症候群、肝炎、血色素症、ウィルソン病、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、癌腫を含む肝癌が含まれ、また自己免疫/炎症疾患も含まれ、その中には後天性免疫不全症候群 (AIDS)、アジソン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫多発性内分泌腺障害症 (APCED) 40
 、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、心筋または心膜の炎症、骨関節炎、骨粗しょう症、膵炎、多発性筋炎、乾癬、ライター症候群、リウマチ様関節炎、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性紅斑性狼瘡、全身性硬化症、血小板減少症、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌の合併症、血液透析、体外循環、ウイルス感染症、細菌感染症、真菌感染症、寄生虫感染症、原虫感染症、蠕虫の感染症及び外傷が含まれ、また代謝障害も含まれ、その中には糖尿病、肥満症及び 50
 骨粗鬆症が含まれ、またアデノウイルス、アレナウイルス、プニヤウイルス、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、パラミクソウイルス、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、ラウドウイルス及びトンガウイルス (tongavirus) として分類されるウイルス性病原体による感染症も含まれる。G C R E C をコードするポリヌクレオチド配列は、サザン法、ノーザン法、ドットプロット法やその他の膜ベースの技術と、P C R 法と、ディップスティック (dipstick) 法、ピン及びマルチフォーマット E L I S A 様アッセイと、変異 G C R E C の発現を検出するために患者から採取した体液または組織を利用するマイクロアレイとにおいて使用し得る。このような定性方法または定量方法は、当分野で公知である。

【0207】

10

或る形態では、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、G C R E C をコードするヌクレオチド配列が有用であり得る。G C R E C をコードするヌクレオチド配列は標準法で標識化され、ハイブリダイゼーション複合体の形成に好適な条件下で、患者から採取した体液または組織のサンプルに添加することができる。好適なインキュベーション期間が経過したらサンプルを洗浄し、シグナルを定量して標準値と比較する。患者サンプルのシグナル量が対照サンプルと比べて著しく変化している場合は、サンプル内の G C R E C をコードするヌクレオチド配列の変異レベルは関連する疾患の存在を示している。このようなアッセイは、動物実験、臨床試験における特定の治療効果を推定するため、或いは個々の患者の治療をモニターするために用いることもできる。

【0208】

20

G C R E C の発現に関連する疾患の診断基準を提供するために、発現のための正常あるいは標準概要を確立する。これは、ハイブリダイゼーション或いは増幅に好適な条件下で、動物或いはヒトの正常な対象から抽出した体液或いは細胞を、G C R E C をコードする配列またはその断片と結合させることにより達成され得る。実質的に精製されたポリヌクレオチドを既知量用いて行った実験から得た値を正常な対象から得た値と比較することにより、標準ハイブリダイゼーションを定量することができる。このようにして得た標準値は、疾患の徴候を示す患者から得たサンプルから得た値と比較することができる。標準値からの偏差を用いて疾患の存在を確定する。

【0209】

30

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

【0210】

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

【0211】

40

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を追加的に診断上利用することは、P C R の利用に参与し得る。これらのオリゴマーは、化学的に合成するか、酵素により生産するか、或いは *in vitro* で産出し得る。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ポリヌクレオチドの断片を含み、最適条件下で特定の遺伝子や条件を識別するべく利用される。また、オリゴマーは、やや緩いストリンジェント条件下で、近縁の D N A 或いは R N A 配列の検出、定量、或いはその両方のため用いることが可能である。

【0212】

或る態様において、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

50

チドプライマーを用いて1塩基性多形性(SNP)を検出し得る。SNPは、多くの場合にヒトの先天性または後天性遺伝病の原因となるような置換、挿入及び欠失である。限定するものではないがSNPの検出方法には、SSCP(single-stranded conformation polymorphism)及び蛍光SSCP(fSSCP)法がある。SSCPでは、GCRC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、ポリメラーゼ連鎖反応法(PCR)を用いたDNAの増幅を行う。DNAは例えば、病変組織または正常組織、生検サンプル、体液その他に由来し得る。DNA内のSNPは、一本鎖形状のPCR生成物の2次及び3次構造に差異を生じさせる。差異は非変性ゲル中でのゲル電気泳動法を用いて検出可能である。fSSCPでは、オリゴヌクレオチドプライマーを蛍光性に標識する。それによってDNAシー

10

【0213】

GCRCの発現を定量するために用い得る方法には、ヌクレオチドの放射標識またはビオチン標識、調節核酸の相互増幅(coamplification)及び標準曲線から得た結果の補間もある(例えば、Melby, P. C.ら(1993) J. Immunol. Methods, 159: 235-244; Duplaa, C.ら(1993) Anal. Biochem. 212: 229-236を参照)。目的のオリゴマーまたはポリヌクレオチドが種々の希釈液中に存在し、分光光度法または比色反応によって定量が迅速になるような高処理フォーマットのアッセイを行うことによって、複数のサンプルの定量速度を加速することができる。

20

【0214】

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由来のオリゴヌクレオチドまたはより長い断片を、マイクロアレイにおけるエレメントとして用いることができる。多数の遺伝子の関連発現レベルを同時にモニターする転写イメージング技術にマイクロアレイを用いることが可能である。これについては、以下に記載する。マイクロアレイはまた、遺伝変異体、突然変異及び多形性の同定に用いることができる。この情報を用いることで、遺伝子機能を決定し、疾患の遺伝的根拠を理解し、疾患を診断し、遺伝子発現の機能としての疾病の進行/後退をモニターし、疾病治療における薬剤の活性を開発及びモニターすることができる。特に、患者にとって最もふさわしく、有効的な治療法を選択するために、この情報を用いて患者の薬理ゲノムプロファイルを開発することができる。例えば、患者の薬理ゲノムプロファイルに基づき、患者に対して高度に有効的で副作用を殆ど示さない治療薬を選択することができる。

30

40

【0215】

別の実施例では、GCRCに特異的な抗体、GCRCまたはその断片をマイクロアレイ上で要素として用いることができる。マイクロアレイを用いて、上記のようなタンパク質-タンパク質相互作用、薬剤-標的相互作用及び遺伝子発現プロファイルをモニターまたは測定することが可能である。

【0216】

或る実施例は、或る組織または細胞タイプの転写イメージを生成するような本発明のポリヌクレオチドの使用に関連する。転写イメージは、特定の組織または細胞タイプにより遺伝子発現の包括的パターンを表す。包括的遺伝子発現パターンは、所与の条件下で所与の時間に発現した遺伝子の数及び相対存在量を定量することにより分析し得る(Seill

50

i a m e r らの米国特許第 5 , 8 4 0 , 4 8 4 号 " C o m p a r a t i v e G e n e T r a n s c r i p t A n a l y s i s " 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タイプの転写または逆転写全体に本発明のポリヌクレオチドまたはその補体をハイブリダイズすることにより、転写イメージを生成し得る。或る実施例では、本発明のポリヌクレオチドまたはその補体がマイクロアレイ上のエレメントのサブセットを複数含むような高処理フォーマットでハイブリダイゼーションを発生させる。結果として得られる転写イメージは、遺伝子活性のプロフィールを提供し得る。

【 0 2 1 7 】

転写イメージは、組織、株化細胞、生検またはその生物学的サンプルから単離した転写物を用いて生成し得る。転写イメージは従って、組織または生検サンプルの場合には i n v i v o、または株化細胞の場合には i n v i t r o での遺伝子発現を反映する。 10

【 0 2 1 8 】

本発明のポリヌクレオチドの発現のプロフィールを作製する転写イメージはまた、工業的または天然の環境化合物の毒性試験のみならず、i n v i t r o モデル系及び薬剤の前臨床評価に関連して使用し得る。全ての化合物は、作用及び毒性のメカニズムを暗示する、しばしば分子フィンガープリントまたは毒性シグネチャと称されるような特徴的な遺伝子発現パターンを惹起する (Nuwaysir, E. F. ら. (1999) Mol. Carcinog. 24:153-159、Steiner, S. 及び N. L. Anderson (2000) Toxicol. Lett. 112-113:467-471、該文献は特に引用することを以って本明細書の一部となす)。試験化合物が、既知の毒性を有する化合物のシグネチャと類似のシグネチャを有する場合には、毒性特性を共有している可能性がある。フィンガープリントまたはシグネチャは、多数の遺伝子及び遺伝子ファミリーからの発現情報を含んでいる場合には、最も有用且つ正確である。理想的には、発現のゲノム全域にわたる測定が最高品質のシグネチャを提供する。自己の発現が任意の試験された化合物により変化しない遺伝子が同様に重要であっても、このような遺伝子の発現レベルを用いて残りの発現データを規準化する。規準化手順は、異なる化合物で処理した後の発現データの比較に有用である。毒性シグネチャの要素に遺伝子機能を割り当てることが毒性メカニズムの解釈に役立つが、毒性の予測につながるシグネチャの統計的に一致させるのに遺伝子機能の知識は必要とされない (例えば 2000 年 2 月 29 日に National Institute of Environmental Health Sciences より発行された Press Release 00-02 を参照されたい。これについては <http://www.niehs.nih.gov/oc/news/toxchip.htm> で入手可能である)。従って、中毒学的スクリーニングの際に毒性シグネチャを用いて、全ての発現した遺伝子配列を含めることは重要且つ望ましいことである。 20 30

【 0 2 1 9 】

或る実施例では、核酸を含有する生物学的サンプルを試験化合物で処理することにより試験化合物の毒性を算定する。処理した生物学的サンプル中で発現した核酸は、本発明のポリヌクレオチドに特異的な 1 つ若しくは複数のプローブでハイブリダイズし、それによって本発明のポリヌクレオチドに対応する転写レベルを定量し得る。処理した生物学的サンプル中の転写レベルを、非処理生物学的サンプル中のレベルと比較する。両サンプルの転写レベルの差は、処理されたサンプル中で試験化合物が引き起こす毒性反応を示す。 40

【 0 2 2 0 】

別の実施例は、本発明のポリペプチド配列を用いて組織または細胞タイプのプロテオームを分析することに関連する。プロテオームの語は、特定の組織または細胞タイプでのタンパク質発現の包括的パターンを指す。プロテオームの各タンパク質成分は、個々に更に分析の対象とすることができる。プロテオーム発現パターン即ちプロフィールは、所与の条件下で所与の時間に発現したタンパク質の数及び相対存在量を定量することにより分析し得る。従って細胞のプロテオームのプロフィールは、特定の組織または細胞タイプのポリ 50

ペプチドを分離及び分析することにより作成し得る。或る実施例では、1次元等電点電気泳動によりサンプルからタンパク質を分離し、2次元ドデシル硫酸ナトリウムスラブゲル電気泳動により分子量に応じて分離するような2次元ゲル電気泳動により分離が達成される(前出のSteiner および Anderson)。タンパク質は、通常クーマシーブルーまたはシルバーまたは蛍光染色などの物質を用いてゲルで染色することにより、分散した、独特な位置にある点としてゲル中で可視化される。各タンパク質スポットの光学密度は、通常サンプル中のタンパク質レベルに比例する。異なるサンプル、例えば試験化合物または治療薬で処理または非処理のいずれかの生物学的サンプルから得られる同等に位置するタンパク質スポットの光学密度を比較し、処理に関連するタンパク質スポット密度の変化を同定する。スポット内のタンパク質は、例えば化学的または酵素的切断とそれに続く質量分析を用いる標準的な方法を用いて部分的にシーケンシングする。スポット内のタンパク質の同一性は、その部分配列を、好適には少なくとも5個の連続するアミノ酸残基を、本発明のポリペプチド配列と比較することにより決定し得る。場合によっては、決定的なタンパク質同定のための更なる配列が得られる。

【0221】

タンパク質の(proteomic)プロフィールは、GCRECに特異的な抗体を用いてGCREC発現レベルを定量することによっても作成し得る。或る実施例では、マイクロアレイ上でエレメントとして抗体を用い、マイクロアレイをサンプルに曝して各アレイ要素へのタンパク質結合レベルを検出することによりタンパク質発現レベルを定量する(Lueking, A.ら(1999) Anal. Biochem. 270:103-111、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、チオールまたはアミノ反応性蛍光化合物を用いてサンプル中のタンパク質を反応させ、各アレイのエレメントにおける蛍光結合の量を検出し得る。

【0222】

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

【0223】

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

【0224】

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

【0225】

マイクロアレイは、本技術分野でよく知られている方法を用いて調製し、使用し、そして

分析する (Brennan, T. M. ら (1995) の米国特許第 5, 474, 796 号、Scheda, M. ら (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619、Baldeschweiler らの (1995) PCT 出願第 WO95/251116 号、Shalon, D. らの (1995) PCT 出願第 WO95/35505 号、Heller, R. A. ら (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155、Heller, M. J. らの (1997) 米国特許第 5, 605, 662 号等を参照)。様々なタイプのマイクロアレイが公知であり、詳細については、DNA Microarrays: A Practical Approach, M. Scheda, 編集。(1999) Oxford University Press, London に記載されている。該文献は、特に引用することを以って本明細書の一部となす。

【0226】

本発明の別の実施例では、天然のゲノム配列をマッピングする際に有効なハイブリダイゼーションプローブを産出するため、GCRC をコードする核酸配列を用いることが可能である。コード配列または非コード配列のいずれかを用いることができ、或る例では、コード配列全体で非コード配列が好ましい。例えば、多重遺伝子ファミリーのメンバー内のコード配列の保存により、染色体マッピング中に望ましくないクロスハイブリダイゼーションが生じる可能性がある。核酸配列は、特定の染色体、染色体の特定領域または人工形成の染色体、例えば、ヒト人工染色体 (HAC)、酵母人工染色体 (YAC)、細菌人工染色体 (BAC)、細菌 P1 産物、或いは単一染色体 cDNA ライブラリに対してマッピングされる (Harrington, J. J. 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を特定の染色体領域の遺伝または制限酵素断片長多型 (RFLP) と関連させるような遺伝子連鎖地図を発生させ得る。(例えば、Lander, E. S. 及び D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357 を参照)。

【0227】

蛍光原位置ハイブリッド形成法 (FISH) は、他の物理的及び遺伝地図データと関連し得る (前出の Heinz-Ulrich, ら (1995) in Meyers, 965-968 ページ、等を参照)。遺伝地図データの例は、種々の科学雑誌あるいは Online Mendelian Inheritance in Man (OMIM) のウェブサイトに見ることができる。物理的染色体地図上の GCRC をコードする遺伝子の位置と特定の疾患との相関性或いは特定の疾患に対する素因は、その疾患に関係する DNA の領域を画定するのに役立つものであり、従って更に位置クローニングする試みとなり得る。

【0228】

確定した染色体マーカーを用いた結合分析等の物理的マッピング技術及び染色体標本原位置ハイブリッド形成法を用いて、遺伝地図を拡張することができる。例えばマウスなど別の哺乳動物の染色体上に遺伝子を配置することにより、正確な染色体の遺伝子座がわかっていない場合でも関連するマーカーを明らかにし得る。この情報は、位置クローニングその他の遺伝子発見技術を用いて疾患遺伝子を探す研究者にとって価値がある。遺伝子または症候群に關与する遺伝子が、血管拡張性失調症の 11q22-23 領域等、特定の遺伝子領域への遺伝的結合によって大まかに位置決めがなされると、該領域に対するいかなる配列マッピングも、更なる調査のための関連遺伝子或いは調節遺伝子を表すことができる (Gatti, R. A. ら (1988) Nature 336:577-580 等を参照)。転座、反転等に起因する、健常者、保有者、感染者の三者間における染色体位置の相違を発見するために本発明のヌクレオチド配列を用いてもよい。

【0229】

本発明の別の実施例では、任意の種々の薬剤スクリーニング技術を以って化合物のライブラリをスクリーニングするために、G C R E C、G C R E Cの触媒作用断片、免疫原断片、またはそのオリゴペプチドを用いることができる。薬剤スクリーニングに用いる断片は、溶液中に遊離しているか、固体支持物に固定されるか、細胞表面上に保持されるか、細胞内に位置することになる。G C R E Cとテストされる薬剤との結合複合の形成は計測できる。

【0230】

別の薬剤スクリーニング方法は、目的のタンパク質に対して好適な結合親和性を有する化合物を高い処理能力でスクリーニングするために用いられる(Gey sen,らの(1984) P C T出願番号 W O 8 4 / 0 3 5 6 4等を参照)。この方法においては、多数の異なる小さな試験用化合物を固体基質上で合成する。試験用化合物は、G C R E C或いはその断片と反応させ、洗浄する。次に、本技術分野でよく知られている方法で、結合したG C R E Cを検出する。精製したG C R E Cはまた、上記した薬剤のスクリーニング技術において用いるプレート上で直接コーティングすることもできる。別法では、非中和抗体を用いてペプチドを捕捉し、ペプチドを固体支持物に固定することもできる。

10

【0231】

別の実施例では、競合薬スクリーニングアッセイを用いることができる。このアッセイでは、G C R E Cを結合することができる中和抗体が、G C R E Cを結合するための試験化合物と特異的に競合する。この方法では、抗体が、1個若しくは数個の抗原決定因子をG C R E Cと共有するペプチドの存在を検出する。

20

【0232】

別の実施例では、新規技術が現在知られているヌクレオチド配列の特性(限定するものではないがトリプレット遺伝暗号及び特異的塩基対の相互作用等を含む)に依存するのであれば、依然として発展すべきいかなる分子生物学技術においても、G C R E Cをコードするヌクレオチド配列を用いることができる。

【0233】

更に詳細説明をしなくとも、当業者であれば以上の説明を以って本発明を最大限に利用できるであろう。従って、これ以下に記載する実施例は単なる例示目的にすぎず、いかようにも本発明を限定するものではない。

【0234】

本明細書において開示した全ての特許、特許出願及び刊行物、特に米国特許第60/212,483号、第60/213,954号、第60/215,209号、第60/216,595号、第60/218,936号、第60/219,154及び第60/220,141は、言及することをもって本明細書の一部となす。

30

【0235】

(実施例)

1 cDNAライブラリの作製

Incyte cDNAは、L I F E S E Q G O L Dデータベース(Incyte Genomics, Palo Alto CA)に記載されたcDNAライブラリに由来するものであり、表4の列5に列記した。ホモジナイズしてグアニジニウムイソチオシアネート溶液に溶解した組織もあり、また、ホモジナイズしてフェノールまたは好適な変性剤の混合液に溶解した組織もある。変性剤の混合液は、例えばフェノールとグアニジニウムイソチオシアネートの単相溶液であるT R I Z O L(Life Technologies)等である。結果として得られた溶解物は、塩化セシウムクッション上で遠心分離するかクロロホルムで抽出した。イソプロパノールか、酢酸ナトリウムとエタノールか、いずれか一方、或いは別の方法を用いて、溶解物からRNAを沈殿させた。

40

【0236】

RNAの純度を高めるため、RNAのフェノールによる抽出及び沈殿を必要な回数繰り返した。場合によっては、DNアーゼでRNAを処理した。殆どのライブラリでは、オリゴd(T)連結常磁性粒子(Promega)、O L I G O T E Xラテックス粒子(QIA

50

GEN, Valencia CA)またはOLIGOTEX mRNA精製キット(QIAGEN)を用いて、ポリ(A+) RNAを単離した。別法では、別のRNA単離キット、例えばPOLY(A)PURE mRNA精製キット(Ambion, Austin TX)を用いて組織溶解物からRNAを直接単離した。

【0237】

場合によってはStratagene社へのRNA提供を行い、対応するcDNAライブラリをStratagene社が作製することもあった。そうでない場合は、UNIZAPベクターシステム(Stratagene)またはSUPERSCRIPTプラスミドシステム(Life Technologies)を用いて本技術分野で公知の推奨方法または類似の方法でcDNAを合成し、cDNAライブラリを作製した(前出のAusubel, 1997, unit 5.1-6.6等を参照)。逆転写は、オリゴd(T)またはランダムプライマーを用いて開始した。合成オリゴヌクレオチドアダプターを二本鎖cDNAに連結反応させ、好適な制限酵素または酵素でcDNAを消化した。殆どのライブラリに対して、cDNAのサイズ(300~1000bp)選択は、SEPHACRYL S1000、SEPHAROSE CL2BまたはSEPHAROSE CL4Bカラムクロマトグラフィー(Amersham Pharmacia Biotech)、或いは調製用アガロースゲル電気泳動法を用いて行った。合成オリゴヌクレオチドアダプターを二本鎖cDNAに連結反応させ、好適な制限酵素または酵素でcDNAを消化した。好適なプラスミドは、例えばPBLUESCRIPTプラスミド(Stratagene)、pSPORT1プラスミド(Life Technologies)、PCDNA2.1プラスミド(Invitrogen, Carlsbad CA)、PBK-CMVプラスミド(Stratagene)あるいはpINCY(Incyte Genomics, Palo Alto CA)またはその誘導体等である。組換えプラスミドは、Stratagene社のXL1-Blue、XL1-BlueMRFまたはSOLLR、或いはLife Technologies社のDH5、DH10BまたはELECTROMAX DH10Bを含むコンピテント大腸菌細胞に形質転換した。

【0238】

2 cDNAクローンの単離

UNIZAPベクターシステム(Stratagene)を用いた*in vivo*切除によって、或いは細胞溶解によって、実施例1のようにして得たプラスミドを宿主細胞から回収した。MagicまたはWIZARD Minipreps DNA精製システム(Promega)、AGTC Miniprep精製キット(Edge Biosystems, Gaithersburg MD)、QIAGEN社のQIAWELL 8 Plasmid、QIAWELL 8 Plus Plasmid及びQIAWELL 8 Ultra Plasmid 精製システム、R.E.A.L. Prep 96プラスミド精製キットの中から少なくとも1つを用いて、プラスミドを精製した。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せずに、4℃で保管した。

【0239】

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

【0240】

3 シークエンシング及び分析

実施例2に記載したようにプラスミドから回収したIncycyte cDNAを、以下に示すようにシークエンシングした。cDNAのシークエンス反応は、標準的方法或いは高処理装置、例えばABI CATALYST 800 サーマルサイクラー(Applied Biosystems, Foster City, CA)を用いて行った。

10

20

30

40

50

Biosystems)またはPTC-200サーマルサイクラー(MJ Research)をHYDRAマイクロディスペンサー(Robbins Scientific)またはMICROLAB 2200(Hamilton)液体転移システムと併用して処理した。cDNAのシークエンス反応は、Amersham Pharmacia Biotech社が提供する試薬またはABIシークエンシングキット、例えばABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit(Applied Biosystems)に与えられた試薬を用いて準備した。cDNAのシークエンス反応の電気泳動的分離及び標識したポリヌクレオチドの検出には、MEGABACE 1000 DNAシークエンシングシステム(Molecular Dynamics)が、標準ABIプロトコル及び塩基対呼び出しソフトウェアを用いるABI PRISM 373または377シークエンシングシステム(Applied Biosystems)が、或いはその他の本技術分野でよく知られている配列解析システムを用いた。cDNA配列内のリーディングフレームは、標準的方法(前出のAusubel, 1997, unit 7.7に概説)を用いて決定した。cDNA配列の幾つかを選択して、実施例8に記載した方法で配列を伸長させた。

10

20

30

40

【0241】

Incyte cDNA配列に由来するポリヌクレオチド配列は、ベクター、リンカー及びポリ(A)配列を除去し、あいまいな塩基対をマスクすることによって有効性を確認した。その際、BLAST、動的プログラミング及び隣接ジヌクレオチド頻度分析に基づくアルゴリズム及びプログラムを用いた。次に、BLAST、FASTA及びBLIMPSに基づくプログラムを用いて、プログラム中の注釈を得るべく、公共のデータベース、例えばGenBankの霊長類及びげっ歯類、哺乳動物、脊椎動物、真核生物のデータベースと、BLOCKS、PRINTS、DOMO、PRODOM及びPFAM等のHidden Markov Model(HMM)ベースのタンパク質ファミリーデータベースの選択に対するIncyte cDNA配列またはその翻訳を問い合わせた(HMMは、遺伝子ファミリーのコンセンサス1次構造を分析する確率的アプローチである。例えば、Eddy, S.R.(1996) Curr. Opin. Struct. Biol. 6:361-365を参照のこと)。問合せは、BLAST、FASTA、BLIMPS及びHMMERに基づくプログラムを用いて行った。Incyte cDNA配列は、完全長のポリヌクレオチド配列を産出するように構築した。或いは、GenBank cDNA、GenBank EST、ステッチされた配列、ストレッチされた配列またはGenscan予測コード配列(実施例4及び5を参照)を用いてIncyte cDNAの集団を完全長まで伸長させた。Phred、Phrap及びConsedに基づくプログラムを用いて構築し、GenMark、BLAST及びFASTAに基づくプログラムを用いてcDNAの集団をオープンリーディングフレームに対してスクリーニングした。対応する完全長ポリペプチド配列を誘導するべく完全長ポリヌクレオチド配列を翻訳した。或いは、本発明のポリペプチドは完全長翻訳ポリペプチドの任意のメチオニン残基で開始し得る。引き続き、GenBankタンパク質データベース(genpept)、SwissProt、BLOCKS、PRINTS、DOMO、PRODOM及びProsite等のデータベース、PFAM等の隠れマルコフモデル(HMM)ベースのタンパク質ファミリーデータベースに対する問合せによって完全長ポリペプチド配列を分析した。完全長ポリヌクレオチド配列はまた、MACDNASIS PROソフトウェア(日立ソフトウェアエンジニアリング, South San Francisco CA)及びLASERGENEソフトウェア(DNA STAR)を用いて分析した。ポリヌクレオチド及びポリペプチド配列アラインメントは、アラインメントした配列と配列の一致率も計算するMEGALIGNマルチシークエンスアラインメントプログラム(DNA STAR)に組み込まれているようなCLUSTALアルゴリズムによって特定されるデフォルトパラメータを用いて生成する。

50

【0242】

Incyte cDNA及び完全長配列の分析及びアセンブリに利用したツール、プログ

ラム及びアルゴリズムの概略と、適用可能な説明、引用文献、閾値パラメータを表7に示す。用いたツール、プログラム及びアルゴリズムを表7の列1に、それらの簡単な説明を列2に示す。列3は好適な引用文献であり、全ての文献はそっくりそのまま引用を以って本明細書の一部となす。適用可能な場合には、列4は2つの配列が一致する強さを評価するために用いたスコア、確率値その他のパラメータを示す（スコアが高ければ高いほどまた確率値が低いければ低いほど2配列間の相同性が高くなる）。

【0243】

完全長ポリヌクレオチド配列およびポリペプチド配列の組み立て及び分析に用いる上記のプログラムは、SEQ ID NO: 18 - 34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。

10

【0244】

4 ゲノムDNAからのコード配列の同定及び編集

Genscan遺伝子同定プログラムを公衆のゲノム配列データベース（例えばgbpri及びgbhtg）に対して実行することにより、推定上のGタンパク質結合受容体を先ず同定する。Genscanは、様々な生物からゲノムDNA配列を分析する汎用遺伝子同定プログラムである（Burge, C. および S. Karlin (1997) J. Mol. Biol. 268: 78 - 94, Burge, C. 及び S. Karlin (1998) Cuff. Opin. Struct. Biol. 8: 346 - 354等を参照）。プログラムは予測エキソンを連結し、メチオニンから停止コドンに及ぶ構築されたcDNA配列を形成する。Genscanの出力は、ポリヌクレオチド及びポリペプチド配列のFASTAデータベースである。Genscanが一度に分析する配列の最大範囲は、30kbに設定した。いずれのGenscan予測cDNA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Gタンパク質結合受容体のためのPFAMモデムに対して問い合わせることにより分析した。Incyte cDNA配列の相同体をGタンパク質結合受容体として注釈を付けてきたことにより、潜在的Gタンパク質結合受容体も同定した。こうして選択されたGenscan予測配列は、次にBLAST分析により公共データベースgenpept及びgbpriと比較した。必要であれば、genpeptからヒットしたトップのBLASTと比較することによりGenscan予測配列を編集し、余分なまたは取り除かれたエキソンなどのGenscanにより予測された配列のエラーを修正する。BLAST分析はまた、任意のIncyte cDNAまたはGenscan予測配列の公共cDNA適用範囲の発見に用いられるので、転写の証拠を提供する。Incyte cDNA適用範囲が利用できる場合には、この情報を用いてGenscan予測配列を修正または確認した。完全長ポリヌクレオチド配列は、実施例3に記載された構築プロセスを用いて、Incyte cDNA配列及び/または公共のcDNA配列でGenscan予測コード配列を構築することにより得た。或いは、完全長ポリヌクレオチド配列は編集または非編集のGenscan予測コード配列に完全に由来する。

20

30

【0245】

5 cDNA配列データを使ったゲノム配列データの構築

40

ステッチ配列 (Stitched Sequence)

部分cDNA配列は、実施例4に記載のGenscan遺伝子同定プログラムにより予測されたエキソンを用いて伸長させた。実施例3に記載されたように構築された部分cDNAは、ゲノムDNAにマッピングし、関連するcDNA及び1つ若しくは複数のゲノム配列から予測されたGenscanエキソンを含むクラスタに分解した。cDNA及びゲノム情報を統合するべくグラフ理論及び動的プログラミングに基づくアルゴリズムを用いて各クラスタを分析し、引き続いて確認、編集または伸長して完全長配列を産出するような潜在的スプライス変異体を生成した。間隔全体の長さがクラスタ中の2以上の配列に存在するような配列を同定し、そのように同定された間隔は推移により等しいと考えられた。例えば、1つのcDNA及び2つのゲノム配列に間隔が存在する場合、3つの間隔は全て

50

等しいと考えられる。このプロセスは、無関係であるが連続したゲノム配列を cDNA 配列により結び合わせて架橋し得る。このようにして同定された区間を、親配列 (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。1種類の親配列に沿って発生した間隔と間隔との連鎖 (cDNA - cDNA またはゲノム配列 - ゲノム配列) は、親の種類を変える連鎖 (cDNA - ゲノム配列) に優先した。結果として得られるステッチ配列は、BLAST 分析により公共データベース genpept 及び gbpr i に翻訳されて比較された。Genscan により予測された不正確なエキソンは、genpept からヒットしたトップの BLAST と比較することにより修正した。必要な場合には、追加 cDNA 配列を用いるかゲノム DNA の検査により配列を更に伸長させた。

10

【0246】

ストレッチ配列 (Stretched Sequence)

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

20

【0247】

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: 18 - 34 を構築するために用いた配列を、BLAST 及びその他 Smith-Waterman アルゴリズムの実施を用いて、Incyte 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: 18 - 34 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。スタンフォード・ヒトゲノムセンター (SHGC)、ホワイトヘッド・ゲノム研究所 (WIGR)、Genethon 等の公的な情報源から入手可能な放射線ハイブリッド及び遺伝地図データを用いて、クラスタ化された配列が前もってマッピングされたかを測定した。マッピングされた配列がクラスタに含まれている結果、個々の配列番号を含めてそのクラスタの全配列が地図上の位置に割り当てられた。

30

【0248】

地図上の位置は、ヒト染色体の範囲または間隔として表される。センチモルガン間隔の地図上の位置は、染色体の p アームの末端に関連して測定する。(センチモルガン (cM) は、染色体マーカー間の組換え頻度に基づく計測単位である。平均して、1 cM は、ヒト中の DNA の 1 メガベース (Mb) にほぼ等しい。尤も、この値は、組換えのホットスポット及びコールドスポットに起因して広範囲に変化する)。cM 距離は、配列が各クラスター内に含まれるような放射線ハイブリッドマーカーに対して境界を提供するような Genethon によってマッピングされた遺伝マーカーに基づく。NCBI 「GeneMap 99」 (<http://www.ncbi.nlm.nih.gov/genemap>) などの一般個人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

【0249】

50

7 ポリペプチド発現の分析

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

【0250】

BLASTに適用する類似のコンピュータ技術を用いて、GenBankやLifeSeq (Incyte Genomics) 等のcDNAデータベースにおいて同一または関連分子を検索する。ノーザン分析は、多数膜系ハイブリダイゼーションよりも非常に速い。更に、特定の同一を厳密な或いは相同的なものとして分類するか否かを決定するため、コンピュータ検索の感度を変更することができる。検索の基準は積スコアであり、次式で定義される。

【0251】

【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

積スコアは、2つの配列間の類似度及び配列が一致する長さの両方を考慮している。積スコアは、0~100の規準化された値であり、次のようにして求める。BLASTスコアにヌクレオチドの配列一致率を乗じ、その積を2つの配列の短い方の長さの5倍で除する。高スコアリングセグメント対(HSP)に一致する各塩基に+5のスコアを割り当て、各不適性塩基対に-4を割り当てることにより、BLASTスコアを計算する。2つの配列は、2以上のHSPを共有し得る(ギャップにより隔離され得る)。2以上のHSPがある場合には、最高BLASTスコアの塩基対を用いて積スコアを計算する。積スコアは、断片的重畳とBLASTアラインメントの質とのバランスを表す。例えば積スコア100は、比較した2つの配列の短い方の長さ全体にわたって100%一致する場合のみ得られる。積スコア70は、一端が100%一致し、70%重畳しているか、他端が88%一致し、100%重畳しているかのいずれかの場合に得られる。積スコア50は、一端が100%一致し、50%重畳しているか、他端が79%一致し、100%重畳しているかのいずれかの場合に得られる。

【0252】

或いは、GCRECをコードするポリヌクレオチド配列を、ポリヌクレオチド配列が派生した組織源に関連して分析した。例えば或る完全長配列は、Incyte cDNA配列(実施例3を参照)と少なくとも一部は重畳するように構築される。各cDNA配列は、ヒト組織から作製されたcDNAライブラリに由来する。各ヒト組織は、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、肝臓、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、または尿管などの生物/組織のカテゴリーの一つに分類される。各カテゴリーのライブラリ数を数えて、全カテゴリーの総ライブラリ数で除する。同様に、各ヒト組織は、以下の疾患/病状カテゴリー即ち癌、細胞系、発達、炎症、神経性、外傷、心血管、鬱血、その他の1つに分類される。各カテゴリーのライブラリ数を数えて、全カテゴリーの総ライブラリ数で除する。演算結果の割合は、GCRECをコードするcDNAの組織特異発現または疾患特異発現を反映する。cDNA配列及びcDNAライブラリ/組織情報については、LIFESEQ GOLDデータベース(Incyte Genomics, Palo Alto CA)で見られる。

【0253】

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もまた、完全長分子の適切な断片から設計したオリゴヌクレオチドプライマーを用いて該断片を伸長させて生成した。一方のプライマーは既知の断

10

20

30

40

50

片の5'伸長を開始するべく合成し、他方のプライマーは既知の断片の3'伸長を開始するべく合成した。開始プライマーの設計は、長さが約22~30ヌクレオチド、GC含有率が50%以上となり、約68~72の温度で標的配列にアニーリングするように、OLIGO 4.06ソフトウェア(National Biosciences)或いは別の適切なプログラムを用いて、cDNAから設計した。ヘアピン構造及びプライマー-プライマー二量体を生ずるようなヌクレオチドの区間は全て回避した。

【0254】

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

10

【0255】

高忠実度の増幅が、当業者によく知られている方法を利用したPCR法によって得られた。PCRは、PTC-200サーマルサイクラー(MJ Research, Inc.)を用いて96穴プレート内で行った。反応混合液は、鋳型DNA及び200nmolの各プライマー、 Mg^{2+} と $(NH_4)_2SO_4$ とβ-メルカプトエタノールを含むバッファー、Taq DNAポリメラーゼ(Amersham Pharmacia Biotech)、ELONGASE酵素(Life Technologies)、Pfu DNAポリメラーゼ(Stratagene)を含む。プライマーの組、PCI AとPCI Bに対して以下のパラメータで増幅を行った。

ステップ 1 94 で3分間
 ステップ 2 94 で15秒間
 ステップ 3 60 で1分間
 ステップ 4 68 で2分間
 ステップ 5 ステップ2、3、及び4を20回繰り返す
 ステップ 6 68 で5分間
 ステップ 7 4 で保存

20

別法では、プライマー対、T7とSK+に対して以下のパラメータで増幅を行った。

ステップ 1 94 で3分間
 ステップ 2 94 で15秒間
 ステップ 3 57 で1分間
 ステップ 4 68 で2分間
 ステップ 5 ステップ2、3、及び4を20回繰り返す
 ステップ 6 68 で5分間
 ステップ 7 4 で保存

30

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

40

【0256】

伸長させたヌクレオチドは、脱塩及び濃縮して384穴プレートに移し、CviJIコラウイルスエンドヌクラーゼ(Molecular Biology Research, Madison WI)を用いて消化し、pUC18ベクター(Amersham Pharmacia Biotech)への再連結反応前に音波処理またはせん断した。シヨットガン・シークエンシングのために、消化したヌクレオチドを低濃度(0.6~0.8%)のアガロースゲル上で分離し、断片を切除し、寒天をAgar ACE(Prom

50

e g a) で消化した。伸長させたクローンを T 4 リガーゼ (New England B i o l a b s , B e v e r l y M A) を用いて p U C 1 8 ベクター (A m e r s h a m P h a r m a c i a B i o t e c h) に再連結し、P f u D N A ポリメラーゼ (S t r a t a g e n e) で処理して制限部位の張出部 (o v e r h a n g) を満たし、大腸菌細胞に形質移入した。形質移入した細胞を選択して抗生物質を含む培地に移し、それぞれのコロニーを切りとって L B / 2 X カルペニシリン培養液の 3 8 4 ウェルプレートに 3 7 で一晩培養した。

【 0 2 5 7 】

細胞を溶解して、T a q D N A ポリメラーゼ (A m e r s h a m P h a r m a c i a B i o t e c h) 及び P f u D N A ポリメラーゼ (S t r a t a g e n e) を用いて以下の手順で D N A を P C R 増幅した。 10

- ステップ 1 9 4 で 3 分間
- ステップ 2 9 4 で 1 5 秒間
- ステップ 3 6 0 で 1 分間
- ステップ 4 7 2 で 2 分間
- ステップ 5 ステップ 2、3、及び 4 を 2 0 回繰り返す
- ステップ 6 7 2 で 5 分間
- ステップ 7 4 で保存

上記したように P I C O G R E E N 試薬 (M o l e c u l a r P r o b e s) で D N A を定量化した。D N A の回収率が低いサンプルは、上記と同一の条件を用いて再増幅した。サンプルは 2 0 % ジメチルスルホキシド (1 : 2 , v / v) で希釈し、D Y E N A M I C エネルギートランスファー シークエンシングプライマー、及び D Y E N A M I C D I R E C T k i t (A m e r s h a m P h a r m a c i a B i o t e c h) または A B I P R I S M B I G D Y E ターミネーターサイクル シークエンシング反応キット (T e r m i n a t o r c y c l e s e q u e n c i n g r e a d y r e a c t i o n k i t) (A p p l i e d B i o s y s t e m s) を用いてシークエンシングした。 20

【 0 2 5 8 】

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

【 0 2 5 9 】

9 個々のハイブリダイゼーションプローブの標識及び使用

S E Q I D N O : 1 8 - 3 4 から導き出されたハイブリダイゼーションプローブを用いて、c D N A、m R N A、またはゲノム D N A をスクリーニングする。約 2 0 塩基対からなるオリゴヌクレオチドの標識について特に記載するが、より大きなヌクレオチド断片に対しても事実上同一の手順が用いられる。オリゴヌクレオチドを、O L I G O 4 . 0 6 ソフトウェア (N a t i o n a l B i o s c i e n c e) のような最新式のソフトウェアを用いてデザインし、5 0 p m o l の各オリゴマーと、2 5 0 μ C i の [³²P] アデノシン三リン酸 (A m e r s h a m P h a r m a c i a B i o t e c h)、) 及び T 4 ポリヌクレオチドキナーゼ (D u P o n t N E N、B o s t o n M A) とを組み合わせて用いることにより標識する。標識したオリゴヌクレオチドは、S E P H A D E X G - 2 5 超細繊維分子サイズ排除デキストラン ビードカラム (A m e r s h a m P h a r m a c i a B i o t e c h) を用いて十分に精製する。A s e I、B g l I I、E c o R I、P s t I、X b a 1 または P v u I I (D u P o n t N E N) のいずれか 1 つのエンドヌクレアーゼで消化されたヒトゲノム D N A の典型的な膜ベースのハイブリダイゼーション解析において、毎分 1 0 ⁷ カウントの標識されたプローブを含むアリコットを用いる。 40

【 0 2 6 0 】

各消化物から得た D N A は、0 . 7 % アガロースゲル上で分画してナイロン膜 (N y t r a n P l u s , S c h l e i c h e r & S c h u e l l , D u r h a m N H) に移 50

す。ハイブリダイゼーションは、40 で16時間行う。非特異的シグナルを除去するため、例えば0.1×クエン酸ナトリウム食塩水及び0.5%ドデシル硫酸ナトリウムに一致する条件下で、プロットを室温で順次洗浄する。オートラジオグラフィまたはそれに代わるイメージング手段を用いてハイブリダイゼーションパターンを視覚化し、比較する。

【0261】

10 マイクロアレイ

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

10

20

【0262】

完全長cDNA、発現配列タグ（EST）、またはその断片またはオリゴマーは、マイクロアレイのエレメントと成り得る。ハイブリダイゼーションに好適な断片またはオリゴマーを、レーザGENEソフトウェア（DNASTAR）等の本技術分野で公知のソフトウェアを用いて選択することが可能である。アレイエレメントは、生物学的サンプル中でポリヌクレオチドを用いてハイブリダイズされる。生物学的サンプル中のポリヌクレオチドは、検出を容易にするために蛍光標識またはその他の分子タグに抱合される。ハイブリダイゼーション後、生物学的サンプルからハイブリダイズされていないヌクレオチドを除去し、蛍光スキャナを用いて各アレイエレメントにおいてハイブリダイゼーションを検出する。或いは、レーザ脱離及び質量スペクトロメトリを用いてもハイブリダイゼーションを検出し得る。マイクロアレイ上のエレメントにハイブリダイズする各ポリヌクレオチドの相補性の度合及び相対存在度は、算定し得る。一実施例におけるマイクロアレイの調整及び使用について、以下に詳述する。

30

【0263】

組織または細胞サンプルの調製

グアニジウムチオシアネート法を用いて組織サンプルから全RNAを単離し、オリゴ（dT）セルロース法を用いてポリ（A）⁺ RNAを精製する。各ポリ（A）⁺ RNAサンプルを、MMLV逆転写酵素、0.05 pg/μlのオリゴ（dT）プライマー（21mer）、1×第一鎖合成バッファー、0.03 unit/μlのRNアーゼ阻害因子、500 μMのdATP、500 μMのdGTP、500 μMのdTTP、40 μMのdCTP、40 μMのdCTP-Cy3（BDS）またはdCTP-Cy5（Amersham Pharmacia Biotech）を用いて逆転写する。逆転写反応は、GEMBRIGHTキット（Incyte）を用いて200 ng ポリ（A）⁺ RNA含有の25体積ml内で行う。特異的対照ポリ（A）⁺ RNAは、非コード酵母ゲノムDNAから *in vitro* 転写により合成する。37°Cで2時間インキュベートした後、各反応サンプル（1つはCy3、もう1つはCy5標識）は、2.5 mlの0.5 M水酸化ナトリウムで処理し、85 で20分間インキュベートし、反応を停止させてRNAを分解させる。サンプルは、2つの連続するCHROMA SPIN 30ゲル濾過スピンカラム（CLONTECH Laboratories, Inc.（CLONTECH）, Palo Alto CA）を用いて精製する。結合後、2つの反応サンプルは、1 mlのグリコ

40

50

ーゲン (1 mg/ml) を用いて析出させたエタノール、60 ml の酢酸ナトリウム及び 300 ml の 100% エタノールである。サンプルは次に、Speed VAC (Savant Instruments Inc., Holbrook NY) を用いて乾燥して仕上げ、14 μ l の 5 x SSC / 0.2% SDS 中で再懸濁する。

【0264】

マイクロアレイの調製

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

10

【0265】

精製したアレイエレメントは、ポリマーコートされたスライドガラス上に固定する。顕微鏡スライドガラス (Corning) は、処理中及び処理後に 0.1% の SDS 及びアセトン中で超音波をかけ、蒸留水で非常に良く洗って洗浄する。スライドガラスは、4% フッ化水素酸 (VWR Scientific Products Corporation (VWR), West Chester PA) 中でエッチングし、蒸留水で広範囲にわたって洗浄し、95% エタノール中で 0.05% アミノプロピルシラン (Sigma) を用いてコーティングする。コーティングしたスライドガラスは、110 のオープンで硬化させる。

20

【0266】

米国特許第 5,807,522 号で説明されている方法を用いて、コーティングしたガラス基板にアレイエレメントを付加する。該特許は、引用を以って本明細書の一部となす。平均濃度が 100 ng/ μ l のアレイエレメント DNA 1 μ l 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。装置はここで、スライド毎に約 5 nl のアレイエレメントサンプルを加える。

【0267】

マイクロアレイには、STRATALINKER UV 架橋剤 (Stratagene) を用いて UV 架橋する。マイクロアレイは、室温において 0.2% SDS で 1 度洗浄し、蒸留水で 3 度洗浄する。リン酸緩衝生理食塩水 (PBS) (Tropix, Inc., Bedford MA) 中の 0.2% カゼイン中において 60 で 30 分間マイクロアレイをインキュベートした後、前に行ったように 0.2% SDS 及び蒸留水で洗浄することにより、非特異結合部位をブロックする。

30

【0268】

ハイブリダイゼーション

ハイブリダイゼーション反応液は、5 x 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 μ g 含む 9 μ l のサンプル混合体を含めたものである。サンプル混合体は、65 まで 5 分間加熱し、マイクロアレイ表面上で等分して 1.8 cm² のカバーガラスで覆う。アレイは、顕微鏡スライドより僅かに大きい空洞を有する防水チェンバーに移す。チェンバーのコーナーに 140 μ l の 5 x SSC を加えることにより、チェンバー内部を湿度 100% に保持する。アレイを含むチェンバーは、60 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 x SSC, 0.1% SDS) において 45 で 10 分間洗浄し、第 2 洗浄緩衝液中 (0.1 x SSC) において 45 で 10 分間各々 3 度洗浄して乾燥させる。

40

【0269】

検出

レポーター標識ハイブリダイゼーション複合体は、Cy3 の励起のためには 488 nm、Cy5 の励起のためには 632 nm でスペクトル線を発生し得る Innova 70 混合

50

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×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の焦点を当てる。アレイを含むスライドを顕微鏡のコンピュータ制御のX-Yステージに置き、対物レンズを通過してラスタースキャンする。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した。

【0270】

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

10

【0271】

スキャンの感度は通常、既知濃度のサンプル混合体に添加されるcDNA対照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配列が含まれ、その位置におけるシグナルの強度をハイブリダイジング種の重量比1:100,000に相関させる。異なる源泉 (例えば試験される細胞及び対照細胞など) からの2つのサンプルを、各々異なる蛍光色素で標識し、他と異なって発現した遺伝子を同定するために単一のアレイにハイブリダイズする場合には、その較正を、2つの蛍光色素で較正するcDNAのサンプルを標識し、ハイブリダイゼーション混合体に各々等量を加えることによつて行う。

20

【0272】

光電子増倍管の出力は、IBMコンパチブルPCコンピュータにインストールされた12ビットRTI-835Hアナログ-デジタル (A/D) 変換ボード (Analog Devices, Inc., Norwood MA) を用いてデジタル化される。デジタル化されたデータは、青色 (低シグナル) から赤色 (高シグナル) までの擬似カラー範囲へのリニア20色変換を用いてシグナル強度がマッピングされたようなイメージとして表示される。データは、定量的にも分析される。2つの異なる蛍光色素を同時に励起及び測定する場合には、各蛍光色素の発光スペクトルを用いて、データは先ず蛍光色素間の光学的クロストーク (発光スペクトルの重なり起因する) を補正する。

30

【0273】

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

【0274】

1.1 相補的ポリヌクレオチド

GCRECをコードする配列或いはその任意の一部に対して相補的配列は、天然のGCRECの発現を検出し、低下させ、または阻害するために用いられる。約15~30塩基対を含むオリゴヌクレオチドの使用について記すが、これより小さな或いは大きな配列の断片の場合でも本質的に同じ方法を用いることができる。Oligo4.06ソフトウェア (National Biosciences) とGCRECをコードする配列とを用いて、適切なオリゴヌクレオチドを設計する。転写を阻害するためには、最も独特な5'配列から相補的オリゴヌクレオチドを設計し、これを用いてプロモーターがコーディング配列に結合するのを阻害する。翻訳を阻害するためには、GCRECをコードする転写物にリボソームが結合しないように相補的オリゴヌクレオチドをデザインする。

40

【0275】

50

1.2 GCRECの発現

GCRECの発現及び精製は、細菌またはウイルスをベースにした発現系を用いて行うことができる。細菌でGCRECを発現するために、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、これらに限定するものではない。組換えベクターを、BL21(DE3)等の好適な細菌宿主に形質転換する。抗生物質耐性をもつ細菌が、イソプロピル-Dチオガラクトピラノシド(IPTG)で誘導されるとGCRECを発現する。真核細胞でのGCRECの発現は、一般にバキュロウイルスとして知られている組換えAutographica californica核多角体病ウイルス(ACMNPV)を昆虫細胞株または哺乳動物細胞株に感染させて行う。バキュロウイルスの非必須のポリヘドリン遺伝子を、相同的組換え、或いは転移プラスミドの仲介に關与する細菌仲介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、強い多角体プロモーターによって高いレベルのcDNAの転写が行われる。組換えバキュロウイルスは、多くの場合はSpodoptera frugiperda(Sf9)昆虫細胞に感染に用いられるが、ヒト肝細胞の感染にも用いられることもある。後者の感染の場合は、バキュロウイルスの更なる遺伝的変更が必要になる。(Engelhard, E. K.ら(1994) Proc. Natl. Acad. Sci. USA 91:3224-3227、Sandig, V.ら(1996) Hum. Gene Ther. 7:1937-1945.等を参照)。

【0276】

殆どの発現系では、融合タンパク質としてGCRECを合成するのに例えばグルタチオンSトランスフェラーゼ(GST)またはペプチドエピトープ標識、例えばFLAGや6-Hisを用いる。これらを用いることにより、未精製細胞溶解物から組換え融合タンパク質の親和性ベースの精製を迅速に1ステップで行うことができる。GSTは日本住血吸虫からの26kDaの酵素であり、タンパク質の活性及び抗原性を維持した状態で、固定化グルタチオン上で融合タンパク質の精製を可能とする(Amersham Pharmacia Biotech)。精製後、GSTの部分特定の開発部位においてGCRECからタンパク質的に切断することが可能である。FLAGは8アミノ酸のペプチドであり、市販されているモノクローナル及びポリクローナル抗FLAG抗体(Eastman Kodak)を用いて免疫親和性精製を可能にする。6ヒスチジン残基が連続して伸長した6-Hisは、金属キレート樹脂(QIAGEN)上での精製を可能にする。タンパク質の発現及び精製の方法は、前出のAusubel(1995)10章、16章に記載されている。これらの方法で精製したGCRECを直接用いて、適用可能な場合には実施例16、17及び18のアッセイを行うことができる。

【0277】

1.3 機能的アッセイ

GCREC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。選り抜きのベクターには、pCMV SPORTプラスミド(Life Technologies)及びpCR3.1プラスミド(Invitrogen)が含まれ、どちらもサイトメガロウイルスプロモーターを有する。リポソーム製剤或いは電気穿孔法を用いて、5~10µgの組換えベクターをヒト細胞株、例えば内皮由来または造血由来の細胞株に一時的に形質移入する。更に、標識タンパク質をコードする配列を含む1~2µg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、形質移入細胞と非形質移入細胞を区別する手段が与えられる。また、標識タンパク質の発現によって、cDNAの組換えベクターからの発現を正確に予想できる。標識タンパク質は、例えば緑色蛍光タンパク質(GFP; Clontech)、CD64またはCD64-GFP融合タンパク質から

選択できる。自動化された、レーザー光学に基づく技術であるフローサイトメトリー (FCM) を用いて、GFP または CD64 - GFP を発現する形質移入された細胞を同定し、その細胞のアポトーシス状態や他の細胞特性を評価する。FCM は、細胞死に先行するか或いは同時に発生する現象を診断する蛍光分子の取込を検出して計量する。このような現象として挙げられるのは、プロピジウムヨウ化物による DNA 染色によって計測される核 DNA 内容物の変化、前方光散乱と 90° 側方光散乱によって計測される細胞サイズと顆粒状性の変化、プロモデオキシウリジンの取込量の低下によって計測される DNA 合成の下方調節、特異抗体との反応性によって計測される細胞表面及び細胞内におけるタンパク質の発現の変化、及び蛍光複合アネキシン V タンパク質の細胞表面への結合によって計測される原形質膜組成の変化とがある。フローサイトメトリー法については、Ormerod, M. G. (1994) Flow Cytometry Oxford, New York, NY. に記述がある。 10

【0278】

遺伝子発現における GCREC の影響は、GCREC をコードする配列と CD64 または CD64 - GFP のいずれかが形質移入された高度に精製された細胞集団を用いて評価することができる。CD64 または CD64 - GFP は、形質転換された細胞表面で発現し、ヒト免疫グロブリン G (IgG) の保存領域と結合する。形質転換細胞と非形質転換細胞は、ヒト IgG または CD64 に対する抗体 (DYNAL, Lake Success, NY) で覆われた磁気ビーズを用いて有効に分離することができる。mRNA は、本技術分野で公知の方法で細胞から精製することができる。GCREC その他の目的の遺伝子をコードする mRNA の発現は、ノーザン分析或いはマイクロアレイ技術で分析することができる。 20

【0279】

1.4 GCREC に特異的な抗体の作製

ポリアクリルアミドゲル電気泳動法 (PAGE; Harrington, M. G. (1990) Methods Enzymol. 182: 488 - 495 等を参照) または他の精製技術を用いて実質上精製された GCREC を用いて、標準プロトコルでウサギを免疫化して抗体を産出する。

【0280】

或いは、レーザ GENE ソフトウェア (DNASTAR) を用いて GCREC アミノ酸配列を解析し、免疫抗原性の高い領域を決定する。そして対応するオリゴペプチドを合成し、このオリゴペプチドを用いて当業者によく知られている方法で抗体を生成する。適切なエピトープ、例えば C 末端付近或いは隣接する親水性領域にあるエピトープの選択については、当分野で公知である (前出の Ausubel, 1995, 11 章等を参照) 。 30

【0281】

通常は、長さ約 15 残基のオリゴペプチドを、Fmoc ケミストリを用いる ABI 431A ペプチドシンセサイザ (Applied Biosystems) を用いて合成し、N-マレイミドベンゾイル-N-ヒドロキシスクシンイミドエステル (MBS) を用いた反応によって KLH (Sigma-Aldrich, St. Louis MO) に結合させて、免疫抗原性を高める (前出の Ausubel, 1995 等を参照)。完全フロイントアジュバントにおいてオリゴペプチド-KLM 複合体を用いてウサギを免疫化する。得られた抗血清の抗ペプチド活性及び抗 GCREC 活性を検査するには、ペプチドまたは GCREC を基質に結合し、1% BSA を用いてブロックし、ウサギ抗血清と反応させて洗浄し、さらに放射性ヨウ素で標識したヤギ抗ウサギ IgG と反応させる。 40

【0282】

1.5 特異的な抗体を用いる天然 GCREC の精製

天然または組換え GCREC を、GCREC 特異抗体を用いた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、抗 GCREC 抗体を活性化クロマトグラフィー用樹脂、例えば CNBr 活性化セファロース (Amers 50

ham Pharmacia Biotech) と共有結合させることにより構築する。結合後に、製造者の使用説明書に従って樹脂をブロックし、洗浄する。

【0283】

GCRECを含む培養液をイムノアフィニティークラムに通し、GCRECを優先的に吸着する条件下(例えば洗浄剤が存在する高イオン強度緩衝液)でカラムを洗浄する。抗体とGCRECの結合を破壊する条件(例えばpH2~3の緩衝液、或いは尿素またはチオシアン酸塩イオン等の高濃度のカオトロブ剤)でカラムを溶出させ、GCRECを回収する。

【0284】

1.6 GCRECと相互作用する分子の同定

GCRECと相互作用する分子は、Gタンパク質などのシグナル伝達に関与する分子以外にもアゴニスト及びアンタゴニストを含み得る。GCRECまた断片を、¹²⁵Iボルトンハンター試薬で標識する。(例えばBolton A. E及びW. M. Hunter (1973) Biochem. J. 133:529-539を参照)。GCRECの断片には、例えば3つの細胞外ループのうち1つ若しくは複数のループ、細胞外N末端領域または第3細胞内ループがある。マルチウェルプレートの穴の中に予め配列しておいた候補分子を、標識したGCRECと共にインキュベートして洗浄し、標識したGCREC複合体を有する全ての穴をアッセイする。GCREC濃度を変えて得たデータを用いて、候補分子とのGCRECの数、親和性及び会合の値を計算する。

10

【0285】

或いは、GCRECと相互作用する分子は、Fields, S.及びO. Song (1989) Nature 340:245-246に記載されているような酵母2ハイブリッドシステムを用いて分析するか、またはMATCHMAKERシステム(Clontech)等の2ハイブリッドシステムに基づく市販のキットを用いて分析する。

20

GCRECはまた、ハイスループットな方法で酵母2ハイブリッドシステムを使用するPATHCALLINGプロセス(CuraGen Corp., New Haven CT)に用いて、遺伝子の2大ライブラリに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. ら(2000)米国特許第6,057,101号)。

【0286】

潜在的GCRECアゴニストまたはアンタゴニストは、実施例17及び18に記載のアッセイを用いてGCREC受容体活性の活性化または阻害を試験し得る。候補分子は、既知のGPCR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の化学ライブラリから選択し得る。

30

【0287】

GCRECとGタンパク質などの細胞内シグナル伝達分子の相互作用を検出する方法は、オルファンGタンパク質結合膜7回貫通受容体から得た内部セグメントまたは細胞質ドメインが、既知のGタンパク質結合膜7回貫通受容体の相似ドメインと交換され、Gタンパク質及びオルファン受容体ドメインに活性化された下流シグナル伝達経路を同定するために用いられるという前提に基づく(Kobilka, B. K. ら(1988) Science 240:1310-1316)。相似方法においては、オルファン受容体のドメインが融合タンパク質の一部としてクローニングされ、結合アッセイで用いて特定のGタンパク質との相互作用を実証し得る。研究によれば、Gタンパク質結合膜7回貫通受容体の第3細胞内ループがGタンパク質相互作用及びシグナル伝達に重要であることが示されてきた(Conklin, B. R. ら(1993) Cell 73:631-641)。例えば、GCRECの第3細胞内ループに対応するDNA断片は、ポリメラーゼ連鎖反応法(PCR)により増幅し、pGEX(Pharmacia Biotech)などの融合ベクターにサブクローニングし得る。構成体は誘導された適切な細菌性宿主に形質転換され、融合タンパク質はグルタチオンセファロース4B(Pharmacia Biotech)アフィニティークロマトグラフィーにより細胞溶解産物から精製する。

40

50

【0288】

in vitro 結合アッセイの場合、Gタンパク質を含む細胞抽出物は、50 mM Tris、pH 7.8、1 mM EGTA、5 mMのMgCl₂、20 mMのCHAPS、20%グリセロール、各10 μgのアプロチニン及びロイペプチンと、20 μlの50 mMフッ化フェニルメチルスルフォニルで抽出することにより準備する。溶解産物は絶えず攪拌しながら氷上で45分間インキュベートし、23,000 g、4 で15分間遠心分離して上澄みを集める。750 μgの細胞抽出物をグルタチオンSトランスフェラーゼ(GST)融合タンパク質ビーズを用いて4 で2時間インキュベートする。GSTビーズは、リン酸緩衝食塩水で5回洗浄する。結合Gサブユニットは、百日咳毒素またはコレラ毒素を用いた [³²P]ADPリボシル化により検出する。SDSサンプル緩衝液(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)を添加することにより反応を停止させる。 [³²P]ADP標識されたタンパク質は、10% SDS-PAGEゲルで分離し、オートラジオグラフを行う。ゲル中で分離したタンパク質はニトロセルロースペーパーに移し、プロット(blotto)(5%脱脂粉乳、50 mMのTris-HCl (pH 8.0)、2 mMのCaCl₂、80 mMのNaCl、0.02%のNaN₃及び0.2%ノニデットP-40)を用いて室温で1時間遮断し、Gサブタイプ選択的抗体(1:500; Calbiochem-Novabiochem)で1.5時間インキュベートする。3洗浄後に、プロットはホースラディッシュペルオキシダーゼ(HRP)共役ヤギ抗ウサギ免疫グロブリン(1:2000, Cappel, Westchester PA)でインキュベートし、化学ルミネセンススペースのECL法(Amersham Corp)で可視化する。

【0289】

1.7 GCREC活性の実証

GCREC活性に対するアッセイは、細胞表面におけるGCRECの発現を測定する。GCRECをコードするcDNAは、好適な哺乳動物細胞株に形質移入する。細胞表面タンパク質は、記載されているようにビオチンで標識する(de la Fuente, M. A. ら(1997) Blood 90:2398-2405)。GCREC特異抗体を用いて免疫沈降を実行し、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)及び免疫プロット技術を用いて免疫沈降サンプルを分析する。標識された免疫沈降と未標識免疫沈降の比は、細胞表面に発現したGCRECの量に比例する。

【0290】

或いは、GCREC活性のためのアッセイは、リガンド/受容体が仲介する細胞増殖の調節のための原型アッセイに基づく。このアッセイでは、スイスマウス3T3細胞におけるDNA合成の速度を測定する。当分野で公知の形質移入方法を用いて、GCRECをコードするポリヌクレオチドを含むプラスミドを静止状態の3T3培養細胞に加える。一過性に形質移入された細胞は次に、放射性DNA前駆分子である [³H]チミジンの存在下でインキュベートする。次に、培養した細胞にGCRECリガンドの変化する量を加える。 [³H]チミジンの酸沈殿可能DNAへの取り込みは、ラジオアイソトープカウンターを用いて適切な時間間隔で測定する。取り込み量は、新たに合成されたDNAの量に正比例する。少なくとも100倍のGCRECリガンド密度範囲に対する投与-反応の1次曲線は、受容体活性を示す。ミリリットル当りの活性の1単位は、50%の反応レベルを産出するGCRECの濃度として画定され、100%であれば [³H]チミジンを酸沈殿可能DNAに最大限取り込むことを表す(McKay, I. 及びI. Leigh, eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York NY, .73ページ)。

【0291】

或いは、GCREC活性のためのアッセイは、GPCRファミリータンパク質がGタンパ

ク質活性化第2メッセンジャーシグナル伝達経路(例えばcAMP; Gaucelin, P. ら(1998) J. Biol. Chem. 273:4990-4996)を調整する能力に基づく。全長GCRECをコードするプラスミドを、当分野で周知の方法を用いて哺乳動物細胞株に形質転換する(例えば、チャイニーズハムスター卵巣細胞株(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。培地においてリガンド有りまたはリガンドなしで細胞を30分間インキュベートし、次に培地を除去し、細胞を1Mの過塩素酸で処理して溶解する。溶解産物中のcAMPレベルは、当分野で既知の方法を用いて放射免疫アッセイにより測定する。リガンドに曝された細胞から得た溶解産物中のcAMPレベルをリガンドなしのものと比較したときの変化は、形質移入された細胞に存在するGCRECの量に比例する。 10

【0292】

イノシトールリン酸レベルの変化を測定するためには、 1×10^5 細胞/穴を含む24穴プレートで細胞を成長させ、イノシトールを含まない培地及び $[^3\text{H}]$ ミオイノシトールを用いて $2 \mu\text{Ci}$ /穴で48時間インキュベートする。培地を除去し、10mMのLiClを含む緩衝液で細胞を洗浄し、その後リガンドを添加する。反応は、過塩素酸の添加により停止する。イノシトールリン酸は、Dowex AG1-X8(Bio-Rad)陰イオン交換樹脂及び液体シンチレーションにより計数された全ての標識されたイノシトールリン酸で抽出及び分離される。リガンドに曝された細胞から得た標識されたイノシトールリン酸のレベルをリガンドなしのものと比較したときの変化は、形質移入された細胞に存在するGCRECの量に比例する。 20

【0293】

18 GCRECリガンドの同定

GCRECは、CHO(チャイニーズハムスター卵巣)またはHEK(ヒト胎児腎臓)293などの真核細胞株において発現する。これらの細胞株は、GPCR発現過程が良好であり、発現したGCRECを下流エフェクターに機能的に結合させることができるような広範囲のGタンパク質を含む。候補リガンドの存在下で発現した受容体の活性化に対し、形質転換した細胞をアッセイする。活性は、cAMPまたは Ca^{2+} などの細胞内第2メッセンジャーの変化により測定する。当分野で公知の標準的な方法を用いるか、活性化受容体によるタンパク質キナーゼCの刺激に反応して発光タンパク質(例えばホタルルシフェラーゼまたは緑色蛍光タンパク質)がプロモーターの転写調節下にあるようなレポーター遺伝子アッセイを用いて直接測定する(Milligan, C. ら(1996) Trends Pharmacol. Sci. 17:235-237)。アッセイ技術は、これら第2メッセンジャーシステムの両方に利用可能であり、アデニリルシクラーゼ活性化Flash Plate Assay(NEN Life Sciences Products)などのマルチウェルプレートフォーマットまたはFluo-4 AM(Molecular Probes)などの蛍光 Ca^{2+} 指示薬でFLIPR蛍光定量的プレート読出しシステム(Molecular Devices)を併用して高処理読出しを可能にする。生理的関連性のある第2メッセンジャー経路が知られていない場合、GCRECは、ホスホリパーゼC及び Ca^{2+} 可動化に関与する経路を介してGCRECのシグナル伝達を通すために、広範囲のGタンパク質と結合することが実証されているようなGタンパク質 $G_{15/16}$ と同時発現し得る(Offerrnanns, S.及びM. I. Simon(1995) J. Biol. Chem. 270:15175-15180)。或いは、GCRECは内因性GPCRが不足しているような操作された酵母系に発現し、それによってOCREC活性化スクリーニングに対してバックグラウンドが存在しない利点を提供し得る。これらの酵母系は、ヒトGPCR及びGタンパク質を内因性酵母フェロモン受容体経路の対応する構成要素に対して置換する。シグナルに対する正常な酵母反応を選択的培地の正の成長またはレポーター遺伝子発現に変換するように、下流シグナル伝達経路も変更する(Broach, J. R及びJ. Thorne(1996) Nature 384(supp.):14-16)。既知のGPCRリガ 30 40 50

ンド及びその他天然の生理活性分子を含む推定上のリガンドに対して受容体をスクリーニングする。組織、生物学的流体及び細胞上澄みから抽出した生物学的抽出物もスクリーニングする。

【0294】

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

10

【0295】

(表の簡単な説明)

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

【0296】

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

【0297】

表3は、予測されるモチーフ及びドメインを含む本発明のポリペプチド配列の構造的特徴を、ポリペプチドの分析に用いるための方法、アルゴリズム及び検索可能なデータベースと共に示す。

20

【0298】

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

【0299】

表5は、本発明のポリヌクレオチドの代表的なcDNAライブラリを示す。

【0300】

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

30

【0301】

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

【0302】

表8は、本発明のポリヌクレオチドの組織特異発現を示す。

【表1】

表1

Incyte プロジェクト ID	Incyte ポリシー番号 ID	ポリシー番号 SEQ ID NO:	Incyte ポリシー番号 ID	ポリシー番号 SEQ ID NO:	Incyte ポリシー番号 ID
1714538	1714538C01	1	1714538C01	18	1714538C01
3406743	3406743C01	2	3406743C01	19	3406743C01
3485895	3485895C01	3	3485895C01	20	3485895C01
7476102	7476102C01	4	7476102C01	21	7476102C01
2432942	2432942C01	5	2432942C01	22	2432942C01
4630911	4630911C01	6	4630911C01	23	4630911C01
7472432	7472432C01	7	7472432C01	24	7472432C01
7474977	7474977C01	8	7474977C01	25	7474977C01
7474848	7474848C01	9	7474848C01	26	7474848C01
7655614	7655614C01	10	7655614C01	27	7655614C01
6792419	6792419C01	11	6792419C01	28	6792419C01
7474790	7474790C01	12	7474790C01	29	7474790C01
7474816	7474816C01	13	7474816C01	30	7474816C01
7476172	7476172C01	14	7476172C01	31	7476172C01
7472141	7472141C01	15	7472141C01	32	7472141C01
7472137	7472137C01	16	7472137C01	33	7472137C01
7477934	7477934C01	17	7477934C01	34	7477934C01

10

20

30

【表2】

表 2 - 1

ポリペプチド SEQ ID NO:	Incyte ポリペプチド ID	GenBank ID NO:	確率 スコア	GenBank 相関係
1	1714538CD1	g2865470	1.50E-217	[ヒト] オルファンGタンパク質結合受容体, GPCR (Tan, C.5(1998) Genomics 52:223-229)
2	3406743CD1	g10946201	0	[ヒト] ニューロメチンU 受容体 1
3	3485895CD1	g1902966	3.40E-16	[マウス種] オキシトシン受容体 (Kubota, Y.5(1996) Mol. Cell Endocrinol. 124:25-32)
		g13517983	0	[ヒト] Gタンパク質結合受容体 91
		g767873	1.10E-47	[ドブネズミ] P2Y 7 受容体 (Tokuyama, Y.5(1995) Biochem. Biophys. Res. Commun. 211:211-218)
4	7476102CD1	g202543	1.50E-12	[ドブネズミ] セロトニン受容体
5	2432942CD1	g6006811	6.80E-73	[ハツカネズミ] 蛇行層受容体 (serpentine receptor)
6	4630911CD1	g13517962	0	[ヒト] 推定Gタンパク質結合受容体
7	7472432CD1	g2231669	5.00E-45	[ヒト] アリジン作動性受容体 P2Y9
		g2736345	4.00E-27	[線虫] G 結合タンパク質との類似性を含む。
8	7474977CD1	g310075	2.00E-151	[ドブネズミ] セロトニン受容体
9	7474848CD1	g5019562	1.10E-37	[ヒト] MAS 関連Gタンパク質結合受容体
		g205314	9.70E-44	[ドブネズミ] mas オルファンGPCR によってコードされるタンパク質 (Young, D.5(1988) Proc. Natl. Acad. Sci. USA 85:5339-5342)
10	7655614CD1	g189270	2.90E-147	[ヒト] 白血球の血小板活性化因子受容体 (GPCR) (Kunz, D.5(1992) J. Biol. Chem. 267:9101-9106)
11	6792419CD1	g4580924	2.30E-289	[ヒト] エンドレジン受容体 B δ 3 (Tsutsumi, M. 5(1999) Gene 228:43-49)
12	7474790CD1	g2198745	2.10E-30	[カイウサギ] α1a-アドレナリン受容体

【表 3】

表 2-2

ポリヌクレオチド SEQ ID NO:	Incyte ポリヌクレオチド ID	GenBank ID NO:	塩率 スコア	GenBank 相同体
13	7474816CD1	g11967419 g1055254 g5809686 g3130153	4.00E-41 2.70E-25 2.60E-163 2.80E-117	[ハツカネズミ] 嗅覚受容体 VIRC3 [ドブネズミ] エロヒン受容体 VN6 (Dulac, C. 及び R. Axel (1995) Cell 83:195-206) [キンギョ] 嗅覚受容体 524 [トラアブ] カルシウム 2+ 検出受容体 (Naito, T. (1998) Proc. Natl. Acad. Sci. USA 95:5178-5181)
14	7476172CD1	g7340541	1.90E-148	[ハツカネズミ] bM332P19.1 (新規な膜 7 回貫通受容体タンパク質、ロドプシンファミリー、嗅覚受容体様) (nm17M1-12)
15	7472141CD1	g1256393 g11692583 g12007423 g1016362	2.00E-59 1.00E-116 2.00E-77 3.60E-69	[ドブネズミ] 麻蓄受容体タンパク質 TB 641 (Thomas, M.B. (1996) Gene 178:1-5) [ハツカネズミ] 嗅覚受容体 M34 [ハツカネズミ] T2 嗅覚受容体 [ドブネズミ] OLI 受容体 (Druetel, G. (1995) Receptors Channels 3:33-40)
16	7472137CD1			
17	7477934CD1			

【表 4】

10

20

30

表 3 - 1

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析手法およびデータベース
1	1714538CD1	426	T366 T397 T282 S360	N7 N27 N41 N151 N196	ロトブリン様 GPCR スーパーファミリー PR00237:P62-L86, T95-L116, F141-V163, H177-S198, V232-G255, V295-M319, R338-R364 オルファンGタンパク質結合受容体 PD027492M24-R89 オルファンGタンパク質結合受容体 PD061371:S357-S426 受容体結合 Gタンパク質 膜貫通 糖蛋白質 リン酸化 リド蛋白質 / グルタミン酸 アミノ酸 PD000009:R87-P195 Gタンパク質結合受容体 DM00013:I63-L371, Y67-R259, Y67-L258, L59-R272 Gタンパク質結合受容体:A147-V163 膜貫通ドメイン:M61-C82, R236-V254 膜7回貫通受容体(ロトブリンファミリー) Tm.1:G77-Y356 Gタンパク質結合受容体 BL00237:F127-P166, F240-C251, R290-D316, G348-R364 Gタンパク質結合受容体シグネチャ:L139-V185 344:R43-W176 シグナル解鎖:M1-A58 膜貫通ドメイン:R45-C65 ロトブリン様 GPCR スーパーファミリー PR00237:V44-C68, M82-S103, Q129-P151, A182-L205	BLIMPS-PRINTS BLAST-PRODOM BLAST-PRODOM BLAST-PRODOM BLAST-DOWD MOTIFS HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILESCAN BLAST-DOMO SPSCAN HMMER BLIMPS-PRINTS
2	3406743CD1	259	S41 S131 S156 S242	N15 N27 N60		

【表 5】

表 3-2

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法およびデータベース
3	3485895CD1	379	S111 R216 S222 S375 T55 S96 S142 S307	N53 N101 N104 N109 N213	Gタンパク質結合受容体 BL00237D:N332-R348 Gタンパク質結合受容体シグネチャ:Y145-I191 ロドプシン様 GPCR スーパーファミリー PR00237:Y70-I94, S103-L124, L147-I169, I83-P204, S233-R256, P275-V299, Y322-R348 受容体結合 Gタンパク質 膜貫通 糖蛋白質リン酸化 リジン蛋白質/バロミチン酸 PD000009:S103-I206 Gタンパク質受容体:T153-I169 Gタンパク質結合受容体 DM00013:Y69-R356, K68-G344, L71-L351 膜貫通ドメイン:I187-I206, M234-Y253, L276-M296, N319-L342 膜7回貫通受容体(ロドプシンファミリー) 7tm_1:G85-Y340 Gタンパク質結合受容体 BL00237:W133-P172, F241-Y252, L270-M296 ロドプシン様:L80-L101 ロドプシン様 GPCR スーパーファミリー PR00237:V128-V150, A164-W185, T220-Y243 Gタンパク質結合受容体 DM00013:S46-G333, L54-E196 シグナル領域:M1-A65 膜貫通ドメイン:L87-I106, P211-C231 ロドプシン様 GPCR スーパーファミリー PR00237B:H85-I106	BLIMPS-BLOCKS PROFILESSCAN BLIMPS-PRINTS BLAST-PRODOM MOTIFS BLAST-DOMO HMMER HMMER-PFAM BLIMPS-BLOCKS MOTIFS BLIMPS-PRINTS BLAST-DOMO SPSCAN HMMER BLIMPS-PRINTS
4	7476102CD1	396	S337 Y86	N34		

【表 6】

表 3-3

SEQ ID NO.	Incyte ポリペプチド ID	アミノ酸残基	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析手法およびデータベース
5	2432942CD1	528	S277 S29 S375 S449 S45 S505 S510 T23 T437 T93	N146 N147 N173 N179 N18 N394 N400 N58 N65	Gタンパク質結合受容体 ; PD000752;E243-K515 Gタンパク質結合受容体 EL00649; A:G395-L422; B:G265-L300; C:C314-L339; D:G362-I386; G:S493-I518 セクレチン様 GPCR スーパーファミリ- ER00249; AY250-R274; B:R282-M306; CA316-L339; D:V355-V380; F:D455-F475; G:L485-Q508 シグナルペプチド:M1-T21 膜貫通ドメイン:I251-R274, W470-F497 膜7回貫通受容体(セクレチンファミリ-) 7tm_2:L245-AS12 Latrophilin/CL-1様 GPS ドメイン:GPS:Y185-R238, 推定プリン作用性受容体 P2Y10: PD055638;S16-V83 Gタンパク質結合受容体 EL00237; AW112-F151; B:F222-T233; CG251-S277; D:N312-R328 ロドプシン様 GPCR スーパーファミリ- PR00237:A:I49-Y73; B:A82-R103; C:K126-L148; D:Y161-L182; E:M214-V237; FA256-L280; G:H302-R328 膜貫通ドメイン:A51-Y73, A82-Y106, I134-F152, I165-L183, V211-T233 膜7回貫通受容体(ロドプシンファミリ-);7tm_1:G64-Y320	BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PRINTS HMMER HMMER HMMER-PFAM HMMER-PFAM
6	4630911CD1	361	S141 S239 S313 S323 S332 T185 T215 T233 T41 T79	N192 N350 N39		BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PRINTS HMMER HMMER-PFAM

【表 7】

表 3-4

SEQ ID NO.	Incyte ポリペプチド ID	アミノ酸残基	潜在的リン酸化部位	潜在的グリコシル化部位	シグナチャ配列、ドメインおよびモチーフ	分析 方法および データベース
7	7472432CD1	469	S207 S251 S264 S458 S217 S27 S289 T116 T163	N168 N357 N94	膜貫通ドメイン:M74-W93, I221-L241, L360-M376 Gタンパク質結合受容体: DM00013 P31387 46-367:P46-T369 5-ヒドロキシトリプトミン 5B 受容体: PD027821:M1-H84	HMMER BLAST-DOMO BLAST- PRODOM BLIMPS- BLOCKS
8	7474977CD1	372			Gタンパク質結合受容体: BL00237A:R120-H159 BL00237B:F222-Y233 BL00237C:Q293-T319 BL00237D:N345-N361 ロドプシン様 GPCR スーパーファミリー: PR00237A:L54-P78 PR00237B:F87-P108 PR00237C:H134-I156 PR00237D:A170-L191 PR00237E:A214-Y237 PR00237F:A298-L322 PR00237G:K335-N361 5-ヒドロキシトリプトミン 5B 受容体: PR00519A:A3-P19 PR00519B:E20-P36 PR00519C:P36-V50 PR00519D:E196-R204 PR00519E:R246-V260 PR00519F:V262-V272 膜貫通ドメイン:F48-L68, M300-L318 膜7回貫通受容体(ロドプシンファミリー):7tm1:R69-Y353 Signal_cleavage::M1-A65	BLIMPS- PRINTS BLIMPS- PRINTS HMMER HMMER-PFAM SPSCAN

【表 8】

10

20

30

表 3-5

SEQ ID NO.	Incyte ポリペプチド ID	アミノ酸残基	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチヤ配列、ドメインおよびモチーフ	分析 方法およびデータベース
9	7474848CD1	330	T125, S284, S161, S173	N271	Gタンパク質結合受容体シグネチヤ:F105-L153 ロドプシン様 GPCR スーパーファミリー: PR00237A:I32-G56; PR00237B:F64-N85; PR00237C:M109-V131; PR00237D:L145- K166; PR00237E:L221-L245; PR00237G:H261-K287 Gタンパク質結合受容体、膜貫通糖運白質: PD013244:F193-E309 Gタンパク質 受容体 (O210.doc):A115-V131 Gタンパク質結合受容体:DM000131:F04201 26- 296:I32-Q288 膜貫通ドメイン:I32-R59; C80-I97; Y113- W133; L145-I162; L191-C211; L218- G236 膜7回貫通受容体(ロドプシンファミリー- 7Em_1):G47-Y279	PROFILESSCAN BLIMPS- PRINTS BLAST- PRODOM MOTIFS BLAST-DMO
10	7655614CD1	494	S146, T266, T303, Y406	N54, N393	Gタンパク質結合受容体:BL00237A:C95-FL34; BL00237B:T30-A41; BL00237C:L216- Q242; BL00237D:N271-K287 ロドプシン様 GPCR スーパーファミリー: PR00237A:A105-V129; PR00237B:T138- A159; PR00237C:S191-I213; PR00237D:R225-L246; PR00237E:S278- C301; PR00237F:T329-L353; PR00237G:S369-S395 Gタンパク質結合受容体:BL00237A:A177- P216; BL00237B:Y286-H297; BL00237C:F324-L350; BL00237D:N379-S395	BLIMPS- BLOCKS BLIMPS- PRINTS BLIMPS- BLOCKS

10

20

30

表 3-6

SEQ ID NO:	Incyte 株式会社 ID	アミノ酸残基	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析手法およびデータベース
10					ロドプシン GTP-結合 PD062404; A249-A330 Gタンパク質結合受容体: DK00013 P47211 27-319; L108-N393 膜貫通ドメイン: L110-V129 膜7回貫通受容体 (ロドプシニアミリー); 7tm_1): G120-Y387 Gタンパク質受容体 (0210.pdocb)V197-I213 Gタンパク質結合受容体シグネチャ: F189-G239	BLAST- PRODOM BLAST-DMO HMMER HMMER-PFAM MOTIFS PROFILESCAN
11	6792419CD1	532	S14 S25 S36 S395 S43 S480 S497 S503 S526 T189 T308 T361 Y520	N149 N2 N209	膜貫通ドメイン: T189-I206, V313-M335, W365-M390 膜7回貫通受容体 (ロドプシニアミリー): G208-L476 Gタンパク質受容体: BL00237A: W257-W296; BL00237B: F372-Y383; BL00237C: K406-S432; BL00237D: N468-K484 インドレリン-B 受容体シグネチャ: PR00571A: Q92-G110; PR00571B: G110-T127; PR00571C: P128-W144; PR00571D: R154-T174; PR00571E: D256-V267; PR00571F: H348-A362; PR00571G: F498-N516; PR00571H: D517-S531 インドレリン受容体シグネチャ: PR00366A: C180-C199; PR00366B: R214-P226; PR00366C: S295-E311; PR00366D: P325-G339; PR00366E: F355-C373; PR00366F: T379-G396; PR00366G: M397-V411	HMMER HMMER-PFAM MOTIFS PROFILESCAN HMMER HMMER-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS BLIMPS- PRINTS

【表 10】

10

20

30

表 3-7

SEQ ID NO.	Incyte ポリペプチド ID	アミノ酸残基	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析 方法およびデータベース
11					ホバシン受容体シグネチャ: PR00358A:D244-E259; PR00358B:K265-S281; PR00358C:V479-C492 Gタンパク質結合受容体シグネチャ:P268-I315 Gタンパク質受容体:I277-V293 受容体 Gタンパク質結合 膜貫通 ETB 糖蛋白 エンドセリン B 前駆体 PD010509:R91-R166 受容体 Gタンパク質結合 膜貫通 糖蛋白 シグナル I NDセリン タイプ ETB PD004497:T167-C221 受容体 Gタンパク質結合 膜貫通 糖蛋白 シグナル I NDセリン タイプ B PD010472:L478-S532 Gタンパク質結合受容体 DM00013 P28088:194-400:E185-L491 DM00013 P214501:74-385:E185-Q492 DM00013 A541261:63-374:E185-Q492 DM00013 P329401:82-398:E185-L481	BLIMPS-PRINTS PROFILES SCAN MOTIFS BLAST-PRODOM BLAST-PRODOM BLAST-PRODOM BLAST-PRODOM BLAST-DOMO
12	7474790CD1	485	S210 S260 S273 S301 S308 S340 S356 S360 S460 T121 T302 T43 T437	N338	シグナルペプチド:M1-A67 シグナルペプチド:M1-A30 膜貫通ドメイン:T13-L33, A83-S103, S168-Y189, I378-W401 膜7回貫通受容体(ロドシニアファミリー):G25-Y431 Gタンパク質 受容体: BL00237A:W74-P113; BL00237B:I178-Y189; BL00237C:C369-L395; BL00237D:Q423-K439	SPSCAN HMMER HMMER HMMER-PFAM BLIMPS-BLOCKS

【表 1 1】

10

20

30

表 3-8

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析 方法およびデータベース
12					ロドプシン様 GPCR スーパーファミリーシグネチャ: PR00237A: I110-Q34; PR00237B: T43-W64; PR00237C: I88-I110; PR00237D: R124-Y145; PR00237E: T170-F193; PR00237F: A374-L398; PR00237G: I413-K439 Gタンパク質結合受容体シグネチャ: L84-L129 Gタンパク質受容体: A94-I110 受容体結合 Gタンパク質 膜貫通 糖蛋白質 糖化 リボ蛋白質 / ミチチン酸タンパク質 アミノ酸 PD000009: Q34-W147 Gタンパク質結合受容体 DM000013: P18130 20-341; P2-Q199, K373-L446 P25100 90-417; A6-R208, K373-L446 P35368 39-364; I3-R198, N338-K440 P50407 76-400; A6-A201, I337-L446 膜貫通ドメイン: Y37-L57 7ドメイン受容体 PD009900: L26-L247 膜貫通ドメイン: I550-I569, F625-A642, Y657-W677, I705-F724 膜7回貫通受容体(代謝調節型) グルタミン酸ファミリー: L548-N799 受容体ファミリー/リガンド結合領域: F15-I456	BLIMPS-PRINTS PROFILES CAN MOTIFS BLAST-PRODOM BLAST-DOMO HMMER BLAST-PRODOM HMMER HMMER-PFAM HMMER-PFAM
13	7474816CD1	255	S124 S130 S2 S303 S333 S636 S841 S879 T168 T169 T214 T407 T425 T461 T55 T613 T66	N145		
14	7476172CD1	881		N212 N257 N285 N331 N38 N39 N405 N522 N545 N688 N74 N832 N865		

10

20

30

【表 1 2】

表 3-9

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析 方法およびデータベース
14					代謝調節型カルシウム結合 GPCR シグネチャ: PR00248B;V23-N38; PR00248C:I37-C56; PR00248D:V96-Y122; PR00248E:L129-Q148; PR00248F:Q148-I164; PR00248G:I164-F181; PR00248I:G622-F643; PR00248J:P659-A682; PR00248K:Y713-A736; PR00248L:N734-P755 受容体、Gタンパク質結合、フィロモン代謝調節型グルタミン酸	BLIMPS-PRINTS
15	7472141GDI	309	S266 T87 T162 T291	N4	PD001315:L661-N799, I550-K727 受容体 シグナル 前駆体 カルシウム結合 糖蛋白 膜リポセサブユニット イオン性 PD001021:L228-I264, Y377-K433 Gタンパク質結合受容体ファミリー-3 DM000837: P35384:I1-694; D513-T800, A29-T409 I59362:I1-893; D513-T800, A29-D410 P31422:I1-859; C517-S825, I25-K305 P31421:I1-850; H525-S802, I25-I345 腫瘍転移ドメイン:M58-L81 M29-V48 膜7回貫通受容体(ロドプシニアファミリー):G40-Y290 Gタンパク質結合受容体 BL00237A:K89-P128 BL00237D:T282-R298 Gタンパク質結合受容体シグネチャ:F102-I147 Gタンパク質結合受容体シグネチャ:T109-I125	BLAST-PRODOM BLAST-PRODOM BLAST-DOMO HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILESSCAN MOTIFS

10

20

30

【表 1 3】

表 3-1-11

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析 方法およびデータベース
16					嗅覚受容体タンパク質 PD000921:L164-L224	BLAST- PRODOM BLAST-DOMO
17	7477934CD1	326	S186 S20 S235 S289 S36 S65 T190 T191	N153 N5	Gタンパク質結合受容体 DM00013 E30955 18-305:I24-I224 腫7回貫通受容体(ロドプシンファミリー) 7tm_1:E39-Y288 Gタンパク質結合受容体 DM00013 A57069 15-304:F33-L303 受容体嗅覚タンパク質 受容体様 Gタンパク質結合 膜貫通 糖蛋白 多重膜伝子ファミリー PD000921:F166-I244 嗅覚受容体シグネチャタンパク質 受容体様 Gタンパ ク質結合 膜貫通 糖蛋白 多重膜伝子ファミリー PD149621:V245-F307 Gタンパク質結合 受容体 BL00237:R88-F127, L205-Y216, T280-L296 嗅覚受容体シグネチャ PR00245:F236-G251, L272-L283, S289-L303, R57-K78, F175-D189 Gタンパク質結合受容体シグネチャ:F100-G145 膜貫通ドメイン:L29-L46 Gタンパク質受容体:G108-I124	BLAST- PRODOM BLAST- PRODOM BLIMPS- BLOCKS BLIMPS- PRINTS PROFILES SCAN HMMER MOTIFS

【表 1 5】

10

20

30

表 4-1

系列番号 SEQ ID NO:	Incyte ポリアノチド ID	配列 長	選択された 断片	配列 断片	5' 位置	3' 位置
18	1714538CB1	2374	1-1852, 2139-2374, 1-234, 818-1572, 532-1151, 68-691	71687857V1 71688177V1 71688955V1 71687846V1 1714538H1 (UGMCMOT02) GNN:g6855250_00025_002_edit 4516658H1 (SINJNOT03) FL210313_00001	2141 986 1246 1684 4 87 2113 1	2287 1547 1843 2287 289 1365 2374 813
19	3406743CB1	813	1-1168, 321-725			
20	3485895CB1	1542	1-534, 626-883, 1518-1542	71296533V1 71755255V1 60149346D1 GNN:g6479070_edit 71295925V1 GBI:g7684460_edit	926 525 87 1 492 1	1542 1186 518 300 1180 1191
21	7476102CB1	1191	1168-1191, 1-1097			
22	2432942CB1	3360	1-1940, 3316-3360, 2803-2843	2432516H1 (BRAVUNT02) 70572441V1 70254651V1 8116752H1 (TONSDIC01) 71733263V1 71873820V1 55046039J2 6544942F6 (LACDRCN02) 71730213V1 70815928V1 2204242T6 (SPINFET02) 71047664V1 7706542J1 (OTRETUE01) GBI:g8112480_edit 6389031F8 (TYMONT03)	1 1447 997 682 300 2675 131 943 2018 2145 1502 1263 924 245 1 459	238 2041 1498 1314 966 3360 943 2827 2854 2139 1660 1516 563 1125 1158
23	4630911CB1	1660	1-1260			

【表 16】

10

20

30

表 4-2

ホリケレオチド SEQ ID NO:	Incyte ホリケレオチド ID	配列 長	選択された 断片	配列 断片	5' 位置	3' 位置
24	7472432CBI	2603	1-24, 2540-2603, 838-2005	6975661HI (BRAHTRDR04) 55037225J2 7341002HI (COLNDIN02) 7059069HI (BRALNON02) FL7472432CBI_00001 2964391F6 (SCORNOF04) g7137674_edit GBI.g8081257_edit	1657 1 2336 1819 239 1353 1 1	2249 393 2603 2477 1648 1728 1119 1018
25	7474977CBI	1119	1-1119			
26	7474848CBI	1018	1-140, 262-344, 546-571, 992-1018			
35	7655614CBI.comp	2177	1537-2177, 1-39, 857-1072, 372-472	7655614HI (UTREME06) FL7655614_g7406476_g156725 7655614JI (UTREME06)	653 705 1	1240 2177 710
28	6792419CBI	1632	1-234, 1-695, 752-809	6314367I6 (NERPTON03) 6110713F8 (MCLDTXT03) g4580923_CD 70619934VI	811 161 1 1048	1620 968 1632 1624
29	7474790CBI	1458	1-103, 370-1458	GNN.g6759183_000017_002_edit	1	1458
30	7474816CBI	1015	1-489, 564-1015	55078072JI FL7474816_g8080066_000028_g1055254	1 248	596 1015
31	7476172CBI	2781	888-2781, 1-352, 1-393, 687-2781	FL7476172_g7684554_000005_g5809686	1	2781
32	7472141CBI	1267	1-449, 1-938	g7940222_edit	1	1267
33	7472137CBI	1559	1268-1559, 1-1043	FL7472137_g8118970_000006_g1256393 55033677HI (FLPAX0017)	624 1	1559 877
34	7477934CBI	981	43-92, 909-981	GBI.g8568403_000012_edit	1	981

【表 17】

10

20

30

表5

ポリスクラオチド SEQ ID NO:	keyte プロジェクトID	代表的 タイプ
18	1714538CB1	ERSTN02
19	3406743CB1	FROSTUS08
20	3485895CB1	KIDNNT020
22	2432942CB1	BRAVUNT02
23	4630911CB1	SPLNFET02
24	7472432CB1	SCORNOT04
27	7655614CB1	PITUDIR01
28	6792419CB1	NERDTR03

10

20

30

【表 1 8】

表 6-1

ライブラリ	ベクター	ライブラリの説明
BRVUNT02	pSPC0T1	ライブラリは非菌源のアストロサイトの別の集団から単離されたポリA RNAを用いて作製した。RNAはポリA RNA 単離用ライブラリ作製用にプールされた。
BESTNON02	PINCY	このノーマライズされた乳房組織ライブラリは、2人のドナーによる2つのライブラリのプールから620,000の独立したクローンから作製した。開始 RNA は 46 才の白人女性から両側小乳房形成術時に採取した乳房組織(ドナー-A)、および60才の白人女性から両側小乳房形成術時に採取した乳房組織(ドナー-B)から作製された。病理は、双方に正常な乳房腺(NA)及び両側乳腫大(B)を示した。患者の病歴には、乳房の肥大、肥満症、腰部神経痛、膝内腔(A)、そして肩の関節痛、中状腺腫瘍、結腸癌、正常分娩、子宮頸癌(B)が含まれる。家族歴には白内障、多形性関節症、子宮癌、良性高血圧症、高脂血症、アルコール性肝臓病、脳血管疾患、耳聾、尿病(A)及び脳血管事故、アテローム性動脈硬化症、結腸癌、II型糖尿病、高脂血症、抑鬱性疾患、アルツハイマー病が含まれている。このライブラリは、極めて長時間(48 時間/一回)の再アニーリングを用いたことを除いては、Scaresら(1994) PNAS 91:9228-9232, Bonaldo S(1996) Genome Research 6:791 を適応した条件を用いて2回にわたってスクリーニングされた。
KIDNNOT20	PINCY	ライブラリは、43 才白人男性の腎臓管切除術、両腎片摘出後に、両腎片摘出時に切除された左腎組織から単離された RNA を用いて作製された。関連する腫瘍組織の病理学検査では、グレード 2 の腎細胞癌が見られた。家族の病歴には、アテローム性動脈硬化症が含まれる。
PITUDIR01	PCDMA2.1	このランダムライブラリ法で得られたライブラリは、顕微鏡性腺癌で死亡した 70 才女性から抽出した下垂体組織から単離した RNA から作製された。腫瘍の病理学検査では、軽度のアルツハイマー病及び軽い悪性腫瘍と脳のアテローム性動脈硬化症を示した。大脳半球、前葉、側頭葉、白質、海馬には、左右に軽度の萎縮が見られた。おびただしい神経原線維変化、老人斑、大脳新皮質領域の大部分に沈着した広汎なアミロイド斑があった。
SPINFET02	PINCY	ライブラリは妊婦 23 週で死亡した白人男子胎児から採取した脾臓組織から単離した RNA を用いて作製された。

【表 19】

表 6-2

ライブラリ	ペグリー	ライブラリの説明
NERDTR03	P1CY	この標準化された後脳神経節細胞のライブラリは、後脳神経節細胞のライブラリの 105 万の独立したクローンから作製した。開始 RNA は、急性脳脊髄炎、急性脳脊髄炎、双方胸腺腫、心臓腫瘍、及び悪性リンパ腫 (チチユルキラー細胞タイプ) のため死亡した 32 才の白人男性の頸部脊髄から取り除かれた後脳神経節細胞から作製された。当患者には、原因不明の発熱、倦怠感、疲労、及び胃腸内出血があった。当患者の病歴には、巨細胞ウイルス感染の強い可能性、肝臓腫瘍、脂肪症、脚腫、出血性尿病、甲狀腺出血、呼吸困難、左肺の肺炎、咽頭のチチユルキラー細胞リンパ腫、ヘルペス、及び HIV-1 に由来する病歴がある。以前に受けた手術には、結腸内視鏡、結腸生検、アディドク腫瘍腫瘍抽出術、鼻咽喉、内視鏡検査、生検等がある。当患者の薬物療法には、Diflucan (fluconazole)、Deltasone (prednisone)、hydrocodone、loratab、アルプラゾラム (Alprazolam)、Resacofone、ProMace-Oxylabom、Etoposide、フスプラチン (Fluorouracil)、シタビジン (Cytarabine)、及びチキマタジン (dexmethasone) 等がある。当患者は放射線療法及び複数の輸血を受けた。このライブラリは、極めて長時間 (48 時間/一回) の凍アニーシングによるハイブリダイゼーションを用いて 2 回にわたって標準化した。
PROST0508	PT/T3	このサブトランスクリプションライブラリは、正常な前立腺腺腫ライブラリ (約 44 万 8 千のクローン) の 21 回のサブトランスクリプションハイブリダイゼーションにかけられた。前立腺腺腫ライブラリの 236 万のクローンを用いて作製した。サブトランスクリプションライブラリは、59 才白人男性の前立腺全摘出術および部分切除リンパ節切除術時に切除された前立腺腫瘍から単離された RNA を用いて作製された。病理は腺癌 (Gleason グレード 3+3) 腺癌腫瘍性形成が見られたことを示した。この患者は前立腺特異抗原 (PSA) の上昇も示した。患者の病歴には結腸憩室炎、アスベスト肺、慢性性静脈炎がある。家歴には多発性骨髄腫、高脂血症およびユーマン種腫瘍炎が含まれている。サブトランスクリプションライブラリの条件は Swaroop 5 (1991) NAR 19:1954 および Bonaldo 5 (1990) Genome Research 6:791 の方法に基づいたものである。
SCORN0704	P1CY	ライブラリは、急性脳脊髄炎、急性脳脊髄炎、両側性胸腺腫、心臓腫瘍、及び悪性リンパ腫 (チチユルキラー細胞タイプ) で死亡した 32 才、白人男性の頸部脊髄から取り除かれた後脳神経節細胞から単離された RNA を使って作製された。患者の病歴には、可能性の高いサイトメガロウイルス、肺炎症、肝性硬変および脂肪肝、脾腫大、出血性尿病、脚腫大、出血性尿病、甲狀腺出血およびヘルペスがある。手術には、結腸鏡検査、大腸生検、アディドク腫瘍抽出術、および上咽頭内視鏡検査および生検が含まれ、治療は放射線療法が含まれている。

【表 20】

表 7-1

【表 2 1】

プログラム	説明	引用文献	パラメータ関連値
ABI FACTURA	核酸配列においてベクター配列を除去して、不定の塩基をマスクするプログラム	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	Fast Data Finderは、アミノ酸または核酸配列の比較、注釈付けに有用である。	Applied Biosystems, Foster City, CA. Paracel Inc., Pasadena, CA.	不一致<50%
ABI オートアセンブ ラ	核酸配列を構築するプログラム。	Applied Biosystems, Foster City, CA.	
BLAST	Basic Local Alignment Search Tool (BLAST) はアミノ酸および核酸配列の配列類似性検索に有用であり、BLAST には5つの機能がある: blastp、blastn、blastx、tblastnおよびtblastx。	Altschul, S.F. 他 (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. 他 (1997) (1997) Nucleic Acids Res. 25:3389-3402.	ESTsの確率値=1.0E-8 以下 完全長配列:確率値= 1.0E-10以下
FASTA	問合せ配列と同種の配列群配との類似性を検索する Pearson および Lipman アルゴリズム。FASTAは最低5つの機能からなる: fasta、tfasta、fastx、tfastxおよびssearch。	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. および M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs/ta F. 値=1.0E-6 構築された ESTs:fasta 同一性=95%以上 一致した長さ=200塩基以上 fastx E 値=1.0E-8 以下 完全長配列: fastx スコア=100以上
BLIMPS	配列をBLOCKS、PRINTS、DOMO、PRODOMおよび PFAM データベースの配列と対応させて遺伝子ファミリー、配列相同性および構造的インガンアプリント領域を検索するBLocks IMProved サーチャー。	Henikoff, S及びJ.G. Henikoff (1991) Nucleic Acids Res.19:6565-6572; Henikoff, J.G及びS. Henikoff (1996) Methods Enzymol.266:88-105, Aitwood, T.K. 他(1997) J. Chem. Inf. Comput. Sci. 37:417-424.	確率値= 1.0E-3以下
HMMER	PFAMのようなタンパク質ファミリーコンセンサス配列の隠れマルコフモデル(HMM)に基づいたデータベースに対して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; Somnhammer, E.L.L. 他 (1988) Nucleic Acids Res.26:320-322; Durbin, R. 他(1998) Our World View, in a Nutshell, Cambridge Univ Press, pp. 1-350.	PFAMヒット:確率値=1.0E-3 以下 シグナルペプチドヒット:スコア=0以上

10

20

30

【表 2 2】

表 7 - 2

プログラム	説明	引用文献	パラメータ閾値
ProfileScan	Prositeで定義された配列パターンと一致するタンパク質配列の構造的モチーフおよび配列モチーフを検索するプログラム。	Gribskov, M. 他 (1988) CABIOS 4:61-66; Gribskov, M. 他 (1989) Methods Enzymol.183:146-159; Bairoch, A. 他(1997) Nucleic Acids Res. 25:217-221.	標準化された順のスコアとその特定の Prosite モチーフに対するGCG指定 HIGSI TM 値 通常、スコア=1.4-2.1.
Phred	高い感度と確率で自動配列決定機のエラーを調べ、修正しアルゴリズム。	Ewing, B. 他 (1998) Genome Res. 8:175-185; Ewing, B.及UP. 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Phils Revised Assembly プログラムで、配列相同性の検索やDNA配列の構築に有用である。	Smith, T.F. および M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. および M.S. Waterman (1981) J. Mol. Biol. 147:195-197; Green, P., University of Washington, Seattle, WA.	スコア=120以上 一致した長さ=56以上
Consed	Phrap で構築したものの表示、編集用グラフィカルツール。	Gordon, D. 他 (1998) Genome Res. 8:195-202.	
SPScan	タンパク質配列をスキャンして、分泌シグナルペプチドの存在を調べる重み付けマトリックス解析プログラム。	Nielson, J. 他 (1997) Protein Engineering 10:1-6; Claverie, J.M.及びS. Audic (1997) CABIOS 12:481-489.	スコア=3.5以上
TMAP	加重マトリックスを用いてタンパク質配列での隣接通セグメントの描写および配向を決定するプログラム。	Persson, B. およびP. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. およびP. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	隠れMarkov モデル(HMM)を使ってタンパク質配列の隣接通セグメントを描写し、配向を決定するプログラム。	Sonnhammer, E.L. 他 (1998) Proc.Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow 他、編集、The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	Prositeで定義された配列と一致したパターンをアミノ酸配列を検索するプログラム。	Bairoch, A. 他 (1997) Nucleic Acids Res. 25:217-221. Wisconsin Package Program Manual, 第9版, M51-59頁, Genetics Computer Group, Madison, WI	

10

20

30

表 8

組織	ポリスクロオチド SEQ ID NO:
乳房、脂肪、皮膚	33 -
筋肉、骨、滑膜	+
結合組織	-
脾臓、肝臓、胆嚢	-
脳: 扁桃、視床、海馬、嗅内皮質、Archaeocortex	-
脳: 線条、尾状核、被蓋、齒状核、淡蒼球、無名質、Ralphie magnus	-
腎臓、胎児結腸、小腸、回腸、食道	-
胎児心臓、大動脈、冠動脈	-
胎児肺、成人肺	-
胎盤、前立腺、子宮	-
嗅球	-

10

20

30

【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
27 December 2001 (27.12.2001)

PCT

(10) International Publication Number
WO 01/98351 A2

(51) International Patent Classification: C07K 14/705

(21) International Application Number: PCT/US01/19275

(22) International Filing Date: 15 June 2001 (15.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/212,483	16 June 2000 (16.06.2000)	US
60/213,954	22 June 2000 (22.06.2000)	US
60/215,209	29 June 2000 (29.06.2000)	US
60/216,595	7 July 2000 (07.07.2000)	US
60/218,936	14 July 2000 (14.07.2000)	US
60/219,154	19 July 2000 (19.07.2000)	US
60/220,141	21 July 2000 (21.07.2000)	US

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

(72) Inventors: and

(75) Inventors/Applicants (for US only): LAL, Preeti [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). HAFALIA, April, J., A. [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). NGUYEN, Daniel, B. [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). GANDHI, Ameena, R. [US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025 (US). KALLICK, Deborah, A. [US/US]; 900 Olive Street, Menlo Park, CA 94025 (US). GRIFFIN, Jennifer, A. [US/US]; 35691 Mello Way, Fremont, CA 94555 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). KHAN, Farrah, A. [IN/US]; 333 Esencula Avenue, #221, Mountain View, CA 94040 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). LI, Dyung, Atna, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). TRIBOULEY, Catherine, M. [FR/US]; 1121 Tennessee Street, #5, San Francisco, CA 94107 (US). LI, Yau [CN/US]; 3885 Corrina Way, Palo Alto, CA

94303 (US). WALIA, Narinder, K. [US/US]; 890 Davis Street, #205, San Leandro, CA 94577 (US). GRAUL, Richard [US/US]; 682-29th Avenue, San Francisco, CA 94121 (US). YAO, Monique, G. [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). YANG, Junming [CN/US]; 7125 Bank Lane, San Jose, CA 95129 (US). RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). AU-YOUNG, Janice [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). HERNANDEZ, Roberto [GB/GB]; 160 Pine Tree Avenue, Canterbury Kent, CT2 7SZ, (GB). WALSH, Roderick, T. [IE/GB]; 8 Boundary Court, St. Lawrence Road, Canterbury, Kent CT1 3EZ, (GB). BOROWSKY, Mark, L. [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). THORNTON, Michael [US/US]; 9 Medway Road, Woodside, CA 94062-2612 (US). HE, Ann [CN/US]; 4601 Catalina Drive, San Jose, CA 95129 (US).

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

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

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BI, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

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

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



WO 01/98351 A2

(54) Title: G-PROTEIN COUPLED RECEPTORS

(57) Abstract: The invention provides human G-protein coupled receptors (GCREC) and polynucleotides which identify and encode GCREC. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of GCREC.

WO 01/98351

PCT/US01/19275

G-PROTEIN COUPLED RECEPTORS**TECHNICAL FIELD**

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

BACKGROUND OF THE INVENTION

Signal transduction is the general process by which cells respond to extracellular signals.
Signal transduction across the plasma membrane begins with the binding of a signal molecule, e.g., a
hormone, neurotransmitter, or growth factor, to a cell membrane receptor. The receptor, thus
activated, triggers an intracellular biochemical cascade that ends with the activation of an intracellular
15 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
25 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
30 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
35 of second messengers such as cyclic AMP (cAMP), phospholipase C, and inositol triphosphate, and

WO 01/98351

PCT/US01/19275

the interaction of the activated GPCR with ion channel proteins. (See, e.g., Watson, S. and 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.)

5 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, glutamate (metabotropic effect), acetylcholine (muscarinic effect), and serotonin); chemokines; lipid
10 mediators of inflammation (e.g., prostaglandins and prostanoids, platelet activating factor, and leukotrienes); and peptide hormones (e.g., bombesin, bradykinin, calcitonin, C5a anaphylatoxin, endothelin, follicle-stimulating hormone (FSH), gonadotropin-releasing hormone (GnRH), neurokinin, and thyrotropin-releasing hormone (TRH), and oxytocin). GPCRs which act as receptors for stimuli that have yet to be identified are known as orphan receptors.

15 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
20 transmembrane domains are 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-methyl-D-aspartate/NMDA receptors). (Reviewed in Watson, S. and S. Arkininstall
35 (1994) The G-Protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 7-9, 19-22,

WO 01/98351

PCT/US01/19275

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 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. (1999) Cell Biochem. Biophys. 30:213-242).

10 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 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 patterns not only in taste and olfactory tissue, but also in male reproductive tissue
15
20 (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 hormone that stimulates the secretion of enzymes and ions in the pancreas and small intestine
25 (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 include the EGF module-containing, mucin-like hormone receptor (Emr1) and CD97 receptor
30 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 (1998) J. Leukoc. Biol. 63:271-280).

35 The third GPCR subfamily is the metabotropic glutamate receptor family. Glutamate is the

WO 01/98351

PCT/US01/19275

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 (Parna, 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_3 -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 disorders, and anxiety; muscarinic agonists are used in the treatment of glaucoma and

WO 01/98351

PCT/US01/19275

tachycardia; serotonin 5HT1D 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 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. 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 treatment of osteoporosis (Mannstadt, 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 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 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 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 01/98351

PCT/US01/19275

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," and "GCREC-17." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-17.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-17. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:18-34.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a

WO 01/98351

PCT/US01/19275

biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is
5 transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a
10 naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID
15 NO:18-34, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group
20 consisting of SEQ ID NO:18-34, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous
30 nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous
35 nucleotides.

WO 01/98351

PCT/US01/19275

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment the composition.

WO 01/98351

PCT/US01/19275

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional GCREC, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test

WO 01/98351

PCT/US01/19275

compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:18-34, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

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

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

Table 3 shows structural features of polypeptide sequences of the invention, including

WO 01/98351

PCT/US01/19275

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

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

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

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

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

Table 8 shows tissue-specific expression of polynucleotides of the invention.

DESCRIPTION OF THE INVENTION

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

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

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be 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

WO 01/98351

PCT/US01/19275

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
5 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
10 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
15 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 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,
20 or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent GCREC. 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
25 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 hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide,
30 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.
35 Amplification is generally carried out using polymerase chain reaction (PCR) technologies well

WO 01/98351

PCT/US01/19275

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
5 directly interacting with GCREC or by acting on components of the biological pathway in which GCREC participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind GCREC polypeptides can be prepared using intact polypeptides or using
10 fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize
15 the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures
20 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
25 phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring
30 nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

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

WO 01/98351

PCT/US01/19275

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,
5 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 GCRC or fragments of GCRC may
10 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., Denhard's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated
15 DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and
20 assembled to produce the consensus sequence.

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

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

WO 01/98351

PCT/US01/19275

	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
5	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

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

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

The term "derivative" refers to a chemically modified polynucleotide or polypeptide.

Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

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

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

A "fragment" is a unique portion of GCRC or the polynucleotide encoding GCRC which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

WO 01/98351

PCT/US01/19275

A fragment of SEQ ID NO:18-34 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:18-34, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:18-34 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish
5 SEQ ID NO:18-34 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:18-34 and the region of SEQ ID NO:18-34 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:1-17 is encoded by a fragment of SEQ ID NO:18-34. A fragment of SEQ ID NO:1-17 comprises a region of unique amino acid sequence that specifically identifies
10 SEQ ID NO:1-17. For example, a fragment of SEQ ID NO:1-17 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-17. The precise length of a fragment of SEQ ID NO:1-17 and the region of SEQ ID NO:1-17 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

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

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

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

25 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
30 Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

35 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

WO 01/98351

PCT/US01/19275

Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

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

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

20 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.

25 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.

30 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

WO 01/98351

PCT/US01/19275

site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of
5 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
10 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
15 *Gap x drop-off: 50*
Expect: 10
Word Size: 3
Filter: on

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

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

The term "humanized antibody" refers to an antibody molecule in which the amino acid
30 sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity.
35 Specific hybridization complexes form under permissive annealing conditions and remain hybridized

WO 01/98351

PCT/US01/19275

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

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

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

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_{0t} or R_{0t} 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.

WO 01/98351

PCT/US01/19275

"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.

5 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.

10 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.

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

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

20 "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.

25 "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.

35 "Probe" refers to nucleic acid sequences encoding GCREC, their complements, or fragments

WO 01/98351

PCT/US01/19275

thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes.

“Primers” are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

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

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

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a “mispriming library,” in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user’s specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping

WO 01/98351

PCT/US01/19275

Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

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

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

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

WO 01/98351

PCT/US01/19275

cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

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

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

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

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

A "transcript image" refers to the collective pattern of gene expression by a particular cell
20 type or tissue under given conditions at a given time.

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

A "transgenic organism," as used herein, is any organism, including but not limited to
30 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

WO 01/98351

PCT/US01/19275

in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human G-protein coupled receptors

WO 01/98351

PCT/US01/19275

(GCREC), the polynucleotides encoding GCREC, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections.

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

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

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

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are G-protein coupled receptors. For example, SEQ ID NO:1 is 99% identical to human orphan G protein-coupled receptor; GPC-R (GenBank ID g2865470) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $1.5e-217$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains an orphan G protein-coupled receptor domain as determined by searching for statistically significant matches in the DOMO database using a Blocks IMProved Searcher (BLIMPS) that searches for gene families, sequence

WO 01/98351

PCT/US01/19275

homology, and structural fingerprint regions. (See Table 3.) SEQ ID NO:1 additionally contains a G-protein coupled receptor subclass 7 transmembrane receptor (rhodopsin family) domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS-PRINTS, BLIMPS-BLOCKS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:1 is a G-protein coupled receptor. In an alternative example, SEQ ID NO:8 is 79% identical to rat serotonin receptor (GenBank ID g310075) as determined by BLAST. (See Table 2.) The BLAST probability score is 2.0e-151. SEQ ID NO:8 also contains G-protein coupled receptor domain structure as determined by searching for statistically significant matches in the HMM-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and PROFILESCAN analyses, as well as BLAST comparisons to protein signature sequences in the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:8 is a G-protein coupled receptor. In an alternative example, SEQ ID NO:10 is 92% identical to the human leukocyte platelet-activating factor receptor, a G-protein coupled receptor (GenBank ID g189270), as determined by BLAST. (See Table 2.) The BLAST probability score is 2.9e-147. SEQ ID NO:10 also contains 7-transmembrane receptor domains, characteristic of G-protein coupled receptors, as well as a receptor binding domain, as determined by searching for statistically significant matches in the HMM-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:10 is a G-protein coupled receptor. In an alternative example, SEQ ID NO:11 is 99% identical to human endothelin receptor B delta 3 (GenBank ID g4580924), as determined by BLAST. (See Table 2.) The BLAST probability score is 2.3e-289. SEQ ID NO:11 also contains 7-transmembrane receptor domains as determined by searching for statistically significant matches in the HMM-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:11 is an endothelin receptor. In an alternative example, SEQ ID NO:16 is 50% identical to the rat taste bud receptor protein TB641 (GenBank ID g1256393), as determined by BLAST. (See Table 2.) The BLAST probability score is 2.0e-59. BLAST searches of the PRODOM and DOMO databases also indicate that SEQ ID NO:16 has homology with receptor protein and G-protein coupled receptor domains. SEQ ID NO:16 also contains a rhodopsin family 7-transmembrane receptor domain, as determined by searching for statistically significant matches in the HMM-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:16 is a G-protein coupled receptor. In an alternative example, SEQ ID NO:17 is 41% identical to the rat G-protein coupled receptor OL1 (GenBank ID g1016362), as determined by BLAST. (See Table 2.) The BLAST probability score is

WO 01/98351

PCT/US01/19275

3.6e-69. SEQ ID NO:17 also contains a 7-transmembrane receptor (rhodopsin family) domain as determined by searching for statistically significant matches in the HMM-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:17 is a G-protein coupled receptor of the rhodopsin family. SEQ ID NO:2-7, SEQ ID NO:9, and SEQ ID NO:12-15 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-17 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:18-34 or that distinguish between SEQ ID NO:18-34 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 2432516H1 is the identification number of an Incyte cDNA sequence, and BRAVUNT02 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., 71687857V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, GNN.g6479070.edit is the identification number of a Genscan-predicted coding sequence, with g6479070 being the GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL210313_00001 represents a "stitched" sequence in which 210313 is the identification number of the cluster of sequences to

WO 01/98351

PCT/US01/19275

which the algorithm was applied, and 00001 is the number of the prediction generated by the algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm. For example, FL7655614_g7406476_g156725 is the identification number of a
5 "stretched" sequence, with 7655614 being the Incyte project identification number, g7406476 being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, and g156725 being the GenBank identification number of the nearest GenBank protein homolog. (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
10 sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors
15 which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows tissue-specific expression of polynucleotides of the invention. Column 1 lists groups of tissues which were tested by polymerase chain reaction (PCR) for expression of the polynucleotides. The remaining columns indicate whether a particular polynucleotide was expressed in each tissue group. Detection of a PCR product indicated positive expression, denoted by a "+"
20 sign, while inability to detect a PCR product indicated a lack of expression, denoted by a "-" sign.

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.
25

The invention also encompasses polynucleotides which encode GCREC. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:18-34, which encodes GCREC. The polynucleotide sequences of SEQ ID NO:18-34, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the
30 sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding GCREC. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding GCREC. A particular aspect of the invention encompasses a variant of a
35 polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID

WO 01/98351

PCT/US01/19275

NO:18-34 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:18-34. 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.

5 It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding 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
10 combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring GCREC, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode GCREC and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring GCREC under appropriately
15 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
20 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
25 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
30 NO:18-34 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
35 the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment

WO 01/98351

PCT/US01/19275

of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) *Short Protocols in Molecular Biology*, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) *Molecular Biology and Biotechnology*, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding GCRC may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include

WO 01/98351

PCT/US01/19275

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.

5 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
10 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
15 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
20 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.

25 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
30 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"
35 breeding and rapid molecular evolution. For example, fragments of a single gene containing random

WO 01/98351

PCT/US01/19275

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
5 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
10 example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) *Proteins, Structures and Molecular Properties*, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of GCREC, or any part thereof, may be altered during direct synthesis and/or
15 combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by
20 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,
25 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
30 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
35 may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used.

WO 01/98351

PCT/US01/19275

(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 GCRC and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding 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); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

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 disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol.*

WO 01/98351

PCT/US01/19275

Chem. 264:5503-5509.) When large quantities of GCREC are needed, e.g. for the production of antibodies, vectors which direct high level expression of GCREC may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

5 Yeast expression systems may be used for production of GCREC. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, *supra*; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994) 10 *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 15 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.)

20 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 25 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 30 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 35 into cell lines using expression vectors which may contain viral origins of replication and/or

WO 01/98351

PCT/US01/19275

endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apv* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β glucuronidase and its substrate β -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding 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 expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding GCREC and that express GCREC may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or 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

WO 01/98351

PCT/US01/19275

include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on GCREC is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, 5 c.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and 10 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. 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, 15 and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic 20 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 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 25 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 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, 30 phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct 35 modification and processing of the foreign protein.

WO 01/98351

PCT/US01/19275

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding GCREC may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric GCREC protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of GCREC activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable 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

WO 01/98351

PCT/US01/19275

E. coli. Cells expressing GCREC or cell membrane fractions which contain GCREC are then contacted with a test compound and binding, stimulation, or inhibition of activity of either GCREC or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with GCREC, either in solution or affixed to a solid support, and detecting the binding of GCREC to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

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

In another embodiment, polynucleotides encoding GCREC or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) *Science* 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) *Clin. Invest.* 97:1999-2002; Wagner, K.U. et al. (1997) *Nucleic Acids Res.* 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred

WO 01/98351

PCT/US01/19275

to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding 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).

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

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of GCREC and G-protein coupled receptors. In addition, the expression of GCREC is closely associated with breast tissue, megakaryoblast cells, prostate tumor, dorsal root ganglion tissue, and pituitary gland tissue. Therefore, GCREC 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 GCREC expression or activity, it is desirable to decrease the expression or activity of GCREC. In the treatment of disorders associated with decreased GCREC expression or activity, it is desirable to increase the expression or activity of GCREC.

Therefore, in one embodiment, GCREC or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), 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,

WO 01/98351

PCT/US01/19275

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, 5 retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal 10 familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, cephalotrigeminal 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, peripheral nervous system 15 disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a 20 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, degenerative valvular 25 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 disorder such 30 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, passive congestion of the

WO 01/98351

PCT/US01/19275

- liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha₁-antitrypsin deficiency, 5
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
- 10 autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, 15 erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, 20 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, 25 hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and tongavirus.

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.

- 30 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

WO 01/98351

PCT/US01/19275

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.

5 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.

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

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

20 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.

30 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-Guerin) and *Corynebacterium parvum* are especially preferable.

35 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

WO 01/98351

PCT/US01/19275

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.

5 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
10 Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda,
15 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.)

20 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.
25 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.)

30 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
35 reactive to two non-interfering GCREC epitopes is generally used, but a competitive binding assay

WO 01/98351

PCT/US01/19275

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 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).

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.)

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.)

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.,

WO 01/98351

PCT/US01/19275

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

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

In a further embodiment of the invention, diseases or disorders caused by deficiencies in GCREC are treated by constructing mammalian expression vectors encoding GCREC and introducing these vectors by mechanical means into GCREC-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic 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).

WO 01/98351

PCT/US01/19275

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

5 constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the

10 ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding GCRC from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID

15 TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of

20 these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to GCRC expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding GCRC 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

25 element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for

30 receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant")

35 discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by

WO 01/98351

PCT/US01/19275

reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; 5 Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding GCREC to cells which have one or more genetic abnormalities with respect to the expression of GCREC. The construction and packaging of adenovirus-based vectors are well 10 known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) *Transplantation* 77:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) 15 *Annu. Rev. Nutr.* 19:511-544 and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding 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 20 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. 25 Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 30 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 35 ordinary skill in the art.

WO 01/98351

PCT/US01/19275

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
5 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
10 alphavirus genome in place of the capsid-coding region results in the production of a large number of GCREC-coding RNAs and the synthesis of high levels of GCREC in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will
15 allow the introduction of GCREC into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

20 Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using
25 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, *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
30 RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding GCREC.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by
35 scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA,

WO 01/98351

PCT/US01/19275

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

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding GCRC. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible 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 GCRC. 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 GCRC expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding GCRC may be therapeutically useful, and in the treatment of disorders associated with decreased GCRC expression or activity, a compound which specifically promotes expression of the polynucleotide encoding GCRC 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

WO 01/98351

PCT/US01/19275

commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding GCREC is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding GCREC are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding GCREC. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

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

WO 01/98351

PCT/US01/19275

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

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

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

10 These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, 15 J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active 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.

20 Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising GCREC or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, GCREC or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to 25 transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) *Science* 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration 30 range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example 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 35 determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such

WO 01/98351

PCT/US01/19275

as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind GCREC may be used for the diagnosis of disorders characterized by expression of GCREC, or in assays to monitor patients being treated with GCREC or agonists, antagonists, or inhibitors of GCREC. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for GCREC include methods which utilize the antibody and a label to detect GCREC in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring GCREC, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of GCREC expression. Normal or standard values for GCREC expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to

WO 01/98351

PCT/US01/19275

GCREC under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of GCREC expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for
5 diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding GCREC may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of GCREC may be correlated
10 with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of GCREC, and to monitor regulation of GCREC levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding GCREC or closely related molecules may be used to identify nucleic acid sequences which encode GCREC. The specificity of the probe, whether it is
15 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%
20 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:18-34 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
25 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
30 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
35 adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in

WO 01/98351

PCT/US01/19275

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,

5 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

10 system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial

15 nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic

20 neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure,

25 ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart

30 disease, congenital heart disease, and complications of cardiac transplantation; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis,

WO 01/98351

PCT/US01/19275

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

In a particular aspect, the nucleotide sequences encoding GCRC may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding GCRC 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

WO 01/98351

PCT/US01/19275

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
5 the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of GCREC, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a
10 sequence, or a fragment thereof, encoding GCREC, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values
15 obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from
20 successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance
25 of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding GCREC may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide
30 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. 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
35 encoding GCREC may be used to detect single nucleotide polymorphisms (SNPs). SNPs are

WO 01/98351

PCT/US01/19275

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

Methods which may also be used to quantify the expression of GCRC include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

WO 01/98351

PCT/US01/19275

In another embodiment, GCREC, fragments of GCREC, or antibodies specific for GCREC may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

5 A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number
10 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The
15 resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present
20 invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson
25 (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is
30 not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to
35 prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of

WO 01/98351

PCT/US01/19275

Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

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

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

A proteomic profile may also be generated using antibodies specific for GCREC to quantify the levels of GCREC expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem.

WO 01/98351

PCT/US01/19275

270:103-111; Mendozze, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

5 Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which
10 alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated
15 biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the
20 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
25 sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

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

35 In another embodiment of the invention, nucleic acid sequences encoding GCRC may be

WO 01/98351

PCT/US01/19275

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

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding 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., Gatti, R.A. et al. (1988) *Nature* 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, 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

WO 01/98351

PCT/US01/19275

between GCREC 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 GCREC, or fragments thereof, and washed. Bound GCREC is then detected by methods well known in the art. Purified GCREC can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

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

In additional embodiments, the nucleotide sequences which encode GCREC may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/212,483, U.S. Ser. No. 60/213,954, U.S. Ser. No. 60/215,209, U.S. Ser. No. 60/216,595, U.S. Ser. No. 60/218,936, U.S. Ser. No. 60/219,154, and U.S. Ser. No. 60/220,141, are expressly incorporated by reference herein.

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. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

WO 01/98351

PCT/US01/19275

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was
5 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
10 recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column
15 chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells
20 including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using
25 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
35 scanner (LabSystems Oy, Helsinki, Finland).

WO 01/98351

PCT/US01/19275

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) *Curr. Opin. Struct. Biol.* 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length

WO 01/98351

PCT/US01/19275

polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program
5 (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate
10 references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide
15 and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:18-34. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative G-protein coupled receptors were initially identified by running the Genscan gene
20 identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) *J. Mol. Biol.* 268:78-94, and Burge, C. and S. Karlin (1998) *Curr. Opin. Struct. Biol.* 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of
25 Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode G-protein coupled receptors, the encoded polypeptides were analyzed by querying against PFAM models for G-protein coupled receptors. Potential G-protein coupled receptors were also identified by homology to Incyte cDNA sequences that had been
30 annotated as G-protein coupled receptors. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted
35 sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available,

WO 01/98351

PCT/US01/19275

this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the *genpept* and *gbpri* public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from *genpept*. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The

WO 01/98351

PCT/US01/19275

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.

5 **VI. Chromosomal Mapping of GREC Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:18-34 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:18-34 were assembled into clusters of contiguous and overlapping sequences using
10 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.

15 Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM
20 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.

25 **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.)

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

35

WO 01/98351

PCT/US01/19275

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

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

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

VIII. Extension of GCREC Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was

WO 01/98351

PCT/US01/19275

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
5 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
10 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
15 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
20 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
25 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
30 shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in
35 384-well plates in LB/2x carb liquid media.

WO 01/98351

PCT/US01/19275

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
5 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).

10 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:18-34 are employed to screen cDNAs,
15 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 μ Ci of
[γ -³²P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase
20 (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).

25 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
30 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
35 aforementioned technologies should be uniform and solid with a non-porous surface (Schna (1999),

WO 01/98351

PCT/US01/19275

supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31.)

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

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 µg/µl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 µl 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 using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

Microarray Preparation

WO 01/98351

PCT/US01/19275

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 amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5
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 washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR
10 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 Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average
15 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. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate
20 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 Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample
25 mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash
30 buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines

WO 01/98351

PCT/US01/19275

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
5 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
10 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
15 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
20 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
25 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
30 integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the GCREC-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring GCREC. Although use of
35 oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same

WO 01/98351

PCT/US01/19275

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.

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 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

WO 01/98351

PCT/US01/19275

shown in 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 include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of GCREC on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding GCREC and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding GCREC and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of GCREC Specific Antibodies

GCREC substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

WO 01/98351

PCT/US01/19275

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 antipeptide 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 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.

Alternatively, molecules interacting with GCREC are analyzed using the yeast two-hybrid

WO 01/98351

PCT/US01/19275

system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech). GCREC may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions
5 between 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
10 combinatorial chemical 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
15 the G-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
20 transduction (Conklin, B.R. et al. (1993) Cell 73:631-641). For example, the DNA fragment corresponding to the third intracellular loop of GCREC may be amplified by the polymerase chain reaction (PCR) and subcloned into a fusion vector such as pGEX (Pharmacia Biotech). The construct is transformed into an appropriate bacterial host, induced, and the fusion protein is purified from the cell lysate by glutathione-Sepharose 4B (Pharmacia Biotech) affinity chromatography.

For *in vitro* binding assays, cell extracts containing G proteins are prepared by extraction with 50 mM Tris, pH 7.8, 1 mM EGTA, 5 mM MgCl₂, 20 mM CHAPS, 20% glycerol, 10 µg of both aprotinin and leupeptin, and 20 µl of 50 mM phenylmethylsulfonyl fluoride. The lysate is incubated on ice for 45 min with constant stirring, centrifuged at 23,000 g for 15 min at 4°C, and the supernatant is collected. 750 µg of cell extract is incubated with glutathione S-transferase (GST)
30 fusion protein beads for 2 h at 4°C. The GST beads are washed five times with phosphate-buffered saline. Bound G subunits are detected by [³²P]ADP-ribosylation with pertussis or cholera toxins. The reactions are terminated by the addition of SDS sample buffer (4.6% (w/v) SDS, 10% (v/v) β-mercaptoethanol, 20% (w/v) glycerol, 95.2 mM Tris-HCl, pH 6.8, 0.01% (w/v) bromophenol blue). The [³²P]ADP-labeled proteins are separated on 10% SDS-PAGE gels, and autoradiographed. The
35 separated proteins in these gels are transferred to nitrocellulose paper, blocked with blotto (5% nonfat

WO 01/98351

PCT/US01/19275

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 G α 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,

5 Westchester PA) and visualized by the chemiluminescence-based ECL method (Amersham Corp.).

XVII. Demonstration of GCREC Activity

An assay for GCREC activity measures the expression of GCREC on the cell surface. cDNA encoding GCREC is transfected into an appropriate mammalian cell line. Cell surface proteins are labeled with biotin as described (de la Fuente, M.A. et al. (1997) *Blood* 90:2398-2405).

10 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
15 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.
20 Incorporation of [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%
25 represents maximal incorporation of [³H]thymidine into acid-precipitable DNA (McKay, I. and I. Leigh, eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York NY, p. 73.)

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

WO 01/98351

PCT/US01/19275

perchloric acid. The cAMP levels in the lysate are measured by radioimmunoassay using methods well-known in the art. Changes in the levels of cAMP in the lysate from cells exposed to ligand compared to those without ligand are proportional to the amount of GCREC present in the transfected cells.

5 To measure changes in inositol phosphate levels, the cells are grown in 24-well plates containing 1×10^5 cells/well and incubated with inositol-free media and [^3H]myo-inositol, 2 $\mu\text{Ci}/\text{well}$, 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
10 total labeled inositol phosphates counted by liquid scintillation. Changes in the levels of labeled inositol phosphate from cells exposed to ligand compared to those without ligand are proportional to the amount of GCREC present in the transfected cells.

XVIII. Identification of GCREC Ligands

GCREC is expressed in a eukaryotic cell line such as CHO (Chinese Hamster Ovary) or HEK
15 (Human Embryonic Kidney) 293 which have a good history of GPCR expression and which contain a wide range of G-proteins allowing for functional coupling of the expressed GCREC to downstream effectors. The transformed cells are assayed for activation of the expressed receptors in the presence of candidate ligands. Activity is measured by changes in intracellular second messengers, such as cyclic AMP or Ca^{2+} . These may be measured directly using standard methods well known in the art,
20 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
25 Assay (NEN Life Sciences Products), or fluorescent Ca^{2+} indicators such as Fluo-4 AM (Molecular Probes) in combination with the FLIPR fluorimetric plate reading system (Molecular Devices). In cases where the physiologically relevant second messenger pathway is not known, GCREC may be coexpressed with the G-proteins $G_{\alpha 12/16}$ which have been demonstrated to couple to a wide range of G-proteins (Offermanns, S. and M.I. Simon (1995) J. Biol. Chem. 270:15175-15180), in order to funnel
30 the signal transduction of the GCREC through a pathway involving phospholipase C and Ca^{2+} mobilization. Alternatively, GCREC may be expressed in engineered yeast systems which lack endogenous GPCRs, thus providing the advantage of a null background for GCREC activation screening. These yeast systems substitute a human GPCR and G_{α} protein for the corresponding components of the endogenous yeast pheromone receptor pathway. Downstream signaling pathways
35 are also modified so that the normal yeast response to the signal is converted to positive growth on

WO 01/98351

PCT/US01/19275

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 ligands including known GPCR ligands and other naturally occurring bioactive molecules. Biological extracts from tissues, biological fluids and cell supernatants are also screened.

5

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific
10 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 01/98351

PCT/US01/19275

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
1714538	1	1714538CD1	18	1714538CB1
3406743	2	3406743CD1	19	3406743CB1
3485895	3	3485895CD1	20	3485895CB1
7476102	4	7476102CD1	21	7476102CB1
2432942	5	2432942CD1	22	2432942CB1
4630911	6	4630911CD1	23	4630911CB1
7472432	7	7472432CD1	24	7472432CB1
7474848	8	7474848CD1	25	7474848CB1
7474848	9	7474848CD1	26	7474848CB1
7655614	10	7655614CD1	27	7655614CB1
6792419	11	6792419CD1	28	6792419CB1
7474790	12	7474790CD1	29	7474790CB1
7474816	13	7474816CD1	30	7474816CB1
7476172	14	7476172CD1	31	7476172CB1
7472141	15	7472141CD1	32	7472141CB1
7472137	16	7472137CD1	33	7472137CB1
7477934	17	7477934CD1	34	7477934CB1

WO 01/98351

PCT/US01/19275

Table 2

Polypeptide SEQ ID NO.	Incyte Polypeptide ID	GenBank ID NO.	Probability Score	GenBank Homolog
1	1714538CD1	g2865470	1.50E-217	[Homo sapiens] orphan G protein-coupled receptor; GPC-R (Tan, C. et al. (1998) Genomics 52:223-229)
2	3406743CD1	g10946201 g1302966	0 3.40E-16	[Homo sapiens] neurexin-1 [Mus sp.] oxytocin receptor (Kubota, Y. et al. (1996) Mol. Cell. Endocrinol. 124:25-32)
3	3485895CD1	g13517983 g767873	0 1.10E-47	[Homo sapiens] G-protein coupled receptor 91 [Rattus norvegicus] P2Y purinoceptor (Tokuyama, Y. et al. (1995) Biochem. Biophys. Res. Commun. 211:211-218)
4	7476102CD1	g202543	1.50E-12	[Rattus norvegicus] serotonin receptor
5	2432942CD1	g6006811	6.80E-73	[Mus musculus] serpinin receptor
6	4630911CD1	g13517962 g2231669	0 5.00E-45	[Homo sapiens] putative purinergic receptor [Homo sapiens] purinergic receptor P2Y9
7	7472432CD1	g2736345	4.00E-27	[Caenorhabditis elegans] contains similarity to G-coupled protein
8	7474977CD1	g110075	2.00E-151	[Rattus norvegicus] serotonin receptor
9	7474848CD1	g5019562 g205314	1.10E-37 9.70E-44	[Homo sapiens] MAS-related G protein-coupled receptor [Rattus norvegicus] mas oncogene (GPCR) encoded protein (Young, D. et al. (1988) Proc. Natl. Acad. Sci. USA 85:5339-5342)
10	7655614CD1	g189270	2.90E-147	[Homo sapiens] leukocyte platelet-activating factor receptor (GPCR) (Kunz, D. et al. (1992) J. Biol. Chem. 267:9101-9106)
11	6792419CD1	g4580924	2.30E-289	[Homo sapiens] endothelin receptor B delta 3 (Teussumi, M. et al. (1999) Gene 238:43-49)
12	7474790CD1	g2198745	2.10E-30	[Oryctolagus cuniculus] alpha 1a-adrenoceptor

WO 01/98351

PCT/US01/19275

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
13	7474816CD1	g11967419 g1055254	4.00E-41 2.70E-25	[Mus musculus] vomeronasal receptor VR6 [Rattus norvegicus] Pheromone receptor VR6 (Dulac, C. and R. Axel. (1995) Cell 83:195-206)
14	7476172CD1	g5809686 g3130153	2.60E-163 2.80E-117	[Carassius auratus] odorant receptor 5.24 [Takifugu rubripes] calcium2+ sensing receptor (Maico, T. et al., (1998) Proc. Natl. Acad. Sci. USA 95:5178-5181)
15	7472141CD1	g7340541	1.90E-148	[Mus musculus] hM32P19.1 (novel 7 transmembrane receptor protein, rhodopsin family, olfactory receptor like (mmu7M1-12)
16	7472137CD1	g1256393	2.00E-59	[Rattus norvegicus] taste bud receptor protein TB 641 (Thomas, M.B. et al. (1996) Gene 178:1-5)
17	7477934CD1	g11692583 g12007423 g1016362	1.00E-116 2.00E-77 3.60E-69	[Mus musculus] odorant receptor M34 [Mus musculus] olfactory receptor [Rattus norvegicus] Olf. receptor (Drutel, G. (1995) Receptors Channels 3:33-40)

WO 01/98351

PCT/US01/19275

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	1714538CD1	426	T366 T397 T282 S360	N7 N27 N41 N151 N196	Rhodopsin-like GPCR superfamily PR00237: P62-L46, T95-L116, F141-V163, H177-S198, V232-G255, V295-M219, H338-R364 ORPHAN G PROTEIN-COUPLED RECEPTOR PD027492: M24-K89 ORPHAN G PROTEIN-COUPLED RECEPTOR PD061371: S357-S426 RECEPTOR COUPLED G-PROTEIN TRANSMEMBRANE GLYCOPROTEIN PHOSPHORYLATION LIPOPROTEIN PALMITATE FAMILY PD000009: R87-P195 G-PROTEIN COUPLED RECEPTORS DK00013: I63-L371, Y67-R259, Y67-L258, L59-R272 G Protein Receptor: A147-V163 Transmembrane domain: M61-C82, A236-V254 7 transmembrane receptor (rhodopsin family) 7tm_1: G77-Y356 G-protein coupled receptor BL00237: F127-P166, F240-C251, R290-D316, G348-R364 G-protein coupled receptors signature: L139-V185 G-PROTEIN COUPLED RECEPTORS DM00013 P30559 32-344:R43-W176 Signal cleavage: M1-A58 Transmembrane domain: R45-C65 Rhodopsin-like GPCR super family PR00237: V44-C68, M82-S103, Q159-P151, A182-L205	BLIMPS- PRINTS BLAST- PRODOM BLAST- PRODOM BLAST- PRODOM BLAST- PRODOM BLAST- PRODOM BLAST- DOMO MOTIFS FIMMER FIMMER-PFAM BLIMPS- BLOCKS PROFILINSCAN BLAST- DOMO SPSCAN FIMMER BLIMPS- PRINTS
2	3406743CD1	259	S41 S131 S156 S242	N15 N27 N60		

WO 01/98351

PCT/US01/19275

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
3	3485895CDL	379	S111 Y216 S222 S375 Y55 S96 S142 S307	N53 N101 N104 N109 N213	G-protein coupled receptor EL00237D: N332-R348 G-protein coupled receptors signature: Y145-I191 Rhodopsin-like GPCR super family PR00237: Y70-I94, S103-L124, L147-I169, I83-P204, S233-A256, P275-V299, Y322-R348 RECEPTOR COUPLED G-PROTEIN TRANSMEMBRANE GLYCOPROTEIN PHOSPHORYLATION LIPOPROTEIN PALMITATE PD00009: S103-I206 G Protein Receptor: T153-I169 G-PROTEIN COUPLED RECEPTORS DM00013: Y69-R356, R68-G344, L71-L351 Transmembrane domain: I187-I206, M234-Y253, L276-M296, N319-L342 7 transmembrane receptor (rhodopsin family) 7tm_1: G85-Y340	BLIMPS-BLOCKS PROFLESOR BLIMPS-PRINTS BLAST-PRODOM BLIMPS-PRINTS BLAST-DOMO HMMER
4	7476102CDL	396	S337 Y86	N34	G-protein coupled receptor EL00237: M133-E172, F241-Y252, L270-M296 Leucine Zipper: L80-L101 Rhodopsin-like GPCR superfamily PR00237: V128-V150, A164-W185, T220-Y243 G-PROTEIN COUPLED RECEPTORS DM00013: S46-G333, L54-E196 Signal cleavage: M1-A65 Transmembrane domain: I87-I106, P211-C231 Rhodopsin-like GPCR super family PR00237B: H85-I106	BLIMPS-BLOCKS PRINTS BLAST-DOMO SPSCAN HMMER BLIMPS-PRINTS

WO 01/98351

PCT/US01/19275

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
5	243294JCD1	528	S277 S29 S375 S449 S45 S505 S510 T23 T437 T93	N146 N147 N173 N179 N18 N394 N400 N58 N65	G-protein coupled receptor: PD00752: E243-K515 G-protein coupled receptor BL00649: A: G395-L422; B: G255-L300; C: C314-L339; D: G362-I386; G: S493-I518 Secretin-like GPCR superfamily PR00249: A: Y250-R274; B: R282-K306; C: A316-L339; D: V355-V380; F: S455-F475; G: L485-Q506 Signal peptide: M1-T21 transmembrane domain: I251-R274, W470-F487 7 transmembrane receptor (Secretin family) 7tm.2: L245-A512 Latrophilin/CL-1-like GPS domain: GPS: Y185-A238,	BLAST- BLAST- PRODOM BLIMPS- BLOCKS BLIMPS- PRINTS HMMER HMMER-PPAM HMMER-PPAM BLAST- PRODOM BLIMPS- BLOCKS BLIMPS- PRINTS HMMER HMMER-PPAM
6	463091LCD1	361	S141 S239 S313 S323 S332 W185 T215 T233 T41 T79	N192 N150 N39	POTATIVE PURINERGIC RECEPTOR P2Y10: PD05638: S16-W68 G-protein coupled receptor BL00237: A: W112-F151; B: F222-T233; C: G251-S277; D: N312-E328 Rhodopsin-like GPCR superfamily PR00237: A: I49-Y73; B: A82-R103; C: K126-L148; D: Y161-L182; E: M214-V217; F: A256-L280; G: H302-E328 Transmembrane domain: A51-Y73, A82- Y106, I134-F152, I165-L183, V211- T233 7 transmembrane receptor (rhodopsin family): 7tm.1: G64-Y320	BLAST- PRODOM BLIMPS- BLOCKS BLIMPS- PRINTS HMMER HMMER-PPAM

WO 01/98351

PCT/US01/19275

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
7	747243CD1	469	S207 S251 S264 S458	N168 N357 N94	transmembrane domain: M74-W93, L221-L241, L560-M378 G-PROTEIN COUPLED RECEPTORS: D800013 P31387 46-367: E46-E369 5-HYDROXYTRYPTAMINE 5H RECEPTOR: PR02782L: M1-H84	EMBL RMMER
8	7474577CD1	372	S217 S27 S289 T116 T163		G-protein coupled receptors: RH00237A: R120-H159 RH00237B: F222-Y233 RH00237C: Q293-T319 RH00237D: N345-N361 Rhodopsin-like GPCR superfamily: RH00237A: L54-F78 RH00237B: P87-F108 RH00237C: H134-I156 RH00237D: A170-L191 RH00237E: A214-Y237 RH00237F: A298-I322 RH00237G: K335-N361 5-hydroxytryptamine 5B receptor: PR00519A: A3-F19 PR00519B: E20-F36 PR00519C: F36-V50 PR00519D: E196-R204 PR00519E: R246-V260 PR00519F: V262-V272 transmembrane domain: F48-L68, M200-L318	EMBL RMMER Pfam SFSCAN

WO 01/98351

PCT/US01/19275

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	7474848CD1	330	T125, S284, S161, S173	N271	G-protein coupled receptors signature: F105-L153 Rhodopsin-like GPCR superfamily: PR00237A: I32-G56; PR00237B: F64-N85; PR00237C: M109-V131; PR00237D: L145-K166; PR00237F: L221-L245; PR00237G: S261-K287 G-PROTEIN COUPLED RECEPTOR, TRANSMEMBRANE GLYCOPROTEIN; P0013244; F193-E3109 G-Protein_Receptor (0210.pdoc): A115-V131 G-PROTEIN COUPLED RECEPTORS: DR00013 P04201 I6-296; I12-Q388 Transmembrane domain: I32-N59; C80-I97; Y113-W133; L145-I162; L191-C211; L218-G236 7 transmembrane receptor (rhodopsin family: 7tm_1): G47-Y279 G-protein coupled receptor: BL00237A: C95-P134; BL00237B: T30-A41; BL00237C: L216-Q242; BL00237D: M271-K287 Rhodopsin-like GPCR superfamily: PR00237A: A105-W129; PR00237B: T138-A159; PR00237C: S191-I213; PR00237D: R325-L246; PR00237F: S278-C301; PR00237F: T329-L353; PR00237G: S169-S195 G-protein coupled receptor: BL00237A: A177-P216; BL00237B: Y286-H287; BL00237C: F124-L350; BL00237D: N379-S395	BLIMFS- PRLMFS EXLMFS BLAST- PRODOM MOTIFS BLAST- DOMO HMMER HMMER-PFAM BLIMFS- BLOCKS BLIMFS- BLOCKS
10	7655614CD1	494	S146, T266, T303, Y406	N54, N193		

WO 01/98351

PCT/US01/19275

Table 3 (cont.)

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10					RHOA35IN GTP-BINDING: PD062404; A249-A330 G-PROTEIN COUPLED RECEPTORS: PR00013 F47211 27-319; L108-W293 Transmembrane domain: L110-V129 7 transmembrane receptor (rhodopsin family: 7tm_1); G120-Y387 G protein receptor (G210.pdoc); V197-I213 G-protein coupled receptors signature: F189-G239 transmembrane domain: T189-I206; V113-M335; W365-M390 7 transmembrane receptor (rhodopsin family): G208-L476 G-PROTEIN_RECEPTOR: BL00237A: W257-W296; BL00237B: F378-X383; BL00237C: K406-S432; BL00237D: M468-K484 Endothelin-B receptor signature: PR00571A: Q82-G110; PR00571B: G110-T127; PR00571C: F428-W144; PR00571D: R154-T174; PR00571E: D256-Y267; PR00571F: H348-A362; PR00571G: F498-N516; PR00571H: D517-S531 Endothelin receptor signature: PR00366A: C180-C199; PR00366B: R214-P226; PR00366C: S295-E311; PR00366D: P325-G339; PR00366E: F355-C373; PR00366F: T179-G396; PR00366G: M397-V411	BLAST PRODOM ELAST-DOMO HMMER HMMER-PFAM MOTIFS PROFILESSCAN HMMER HMMER-PFAM BLASTS- BLOCKS BLASTS- PRINTS BLASTS- PRINTS
11	6792413CD1	532	S14 S25 S36 S395 S43 S480 S497 S503 S526 T189 T308 T361 Y520	N149 N2 N209	signature: F189-G239 transmembrane domain: T189-I206; V113-M335; W365-M390 7 transmembrane receptor (rhodopsin family): G208-L476 G-PROTEIN_RECEPTOR: BL00237A: W257-W296; BL00237B: F378-X383; BL00237C: K406-S432; BL00237D: M468-K484 Endothelin-B receptor signature: PR00571A: Q82-G110; PR00571B: G110-T127; PR00571C: F428-W144; PR00571D: R154-T174; PR00571E: D256-Y267; PR00571F: H348-A362; PR00571G: F498-N516; PR00571H: D517-S531 Endothelin receptor signature: PR00366A: C180-C199; PR00366B: R214-P226; PR00366C: S295-E311; PR00366D: P325-G339; PR00366E: F355-C373; PR00366F: T179-G396; PR00366G: M397-V411	BLAST PRODOM ELAST-DOMO HMMER HMMER-PFAM MOTIFS PROFILESSCAN HMMER HMMER-PFAM BLASTS- BLOCKS BLASTS- PRINTS BLASTS- PRINTS

WO 01/98351

PCT/US01/19275

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
11					Bombesin receptor: signature: P000358A: D244-F259; P000358B: K263-S281; P000358C: V479-C492 G-protein coupled receptors signature: F268-L315 G-Protein Receptor: I277-V293 RECEPTOR G PROTEIN COUPLED TRANSMEMBRANE EXT GLYCOPROTEIN ENDOTHELIN B PRECURSOR P0010509: M91-R166 RECEPTOR G PROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN SIGNAL ENDOTHELIN TYPE ETB P0004497: T167-C221 RECEPTOR G PROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN SIGNAL ENDOTHELIN TYPE B P0010472: L478-S532 G-PROTEIN COUPLED RECEPTORS DM00013 P28088 94-400; E185-L431 DM00013 P21450 74-395; E185-C432 DM00013 A54126 63-374; E185-C432 DM00013 P232940 82-398; E185-L491 signal peptide: M1-A67 signal peptide: M1-A30 transmembrane domain: T13-L33, A83-S103, S168-Y189, I378-W401 7 transmembrane receptor (rhodopsin family): G25-Y431 G-PROTEIN RECEPTOR: BL00237A: W74-P113; BL00237B: I178-Y189; BL00237C: C369-L395; BL00237D: Q423-K439	BLIMFS- BLOCKS PRINFS PROFILINSCAN MOTIFS ELAST- PRODOM ELAST- PRODOM ELAST- PRODOM ELAST- DOMO SPSCAN HMMER HMMER HMMER-PPAM BLIMFS- BLOCKS
12	7474790CD1	485	S210 S260 S273 S301 S308 S340 S356 S360 S460 T121 T102 T43 T437	N338		

WO 01/98351

PCT/US01/19275

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
12					Rhodopsin-like GPCR superfamily signatures: F80237R; Q10 Q34; F80237B; T43-864; F80237C; T88-I110; F80237D; Q124-Q145; F80237E; T176-Q199; F80237F; A374-E528; F80237G; T413-K439 G-protein coupled receptors Signatures: L84-L129 G-PROTEIN RECEPTOR: N94-I110 RECEPTOR COUPLED G PROTEIN TRANSMEMBRANE GLYCOPROTEIN PHOSPHORYLATION LIPOPROTEIN PALMITATE PROTEIN FAMILY F802009: Q34-W147 G-PROTEIN COUPLED RECEPTORS DM00013: E48130 20-361; P2-Q199, K373-L446 E25100 90-417; A6-R208, K373-L446 P3368 39-364; I3-R198, N338-K440 P50407 76-400; A6-A201, E337-L446 transmembrane domain: Y37-L57 PHEROMONE RECEPTOR F8009990: L26-L247 transmembrane domains: I550-I569, F625-A542, Y637-W677, E705-E724 7 transmembrane receptor (metabotropic glutamate family): L548-N799 Receptor family ligand binding region: F15-I456	BLAST- PRODOM EMBL- PRINTS EMBLER- PFAM EMBLER- PFAM
13	747481CDD1	255		N145		EMBLER
14	747617ZDD1	881	S124 S130 S2 S303 S333 S636 S841 S879 T168 T169 T214 T407 T425 T461 T55 T613 T66	N212 N257 N285 N331 N38 N39 N405 N522 M545 N688 N74 N832 N865		EMBLER- PFAM

WO 01/98351

PCT/US01/19275

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 Packages
14					Metabotropic glutamate GPCR Signature: PR00248B: V23-N38, PR00248C: I37-C56, PR00248D: V96-Y22, PR00248E: L129-Q148, PR00248F: Q148-I164, PR00248G: I164-F181, PR00248H: G522-F643, PR00248I: P659-A682, PR00248J: F713-A736, PR00248K: N734-F755 RECEPTOR G PROTEIN COUPLED PHEROMONE METABOTROPIC GLUTAMATE PR001315: L561-N799, I550-K727 RECEPTOR SIGNAL PRECURSOR GLUTAMATE GLYCOPROTEIN MEMBRANE LYASE SUBUNIT IONIC PD001021: L28-L264, Y377-K433 G-PROTEIN COUPLED RECEPTORS FAMILY 3 DM00637: P35384 1-894: D513-T800, A29-D409 I59362 1-893: D513-T800, A29-D410 F31422 1-859: C517-S825, L25-K303 F31421 1-850: H525-S802, L25-I345 transmembrane domain: M56-L61 M29-V48 7 transmembrane receptor (rhodopsin family): G40-Y290 G-protein coupled receptor BL00237A: K39-F128 BL00237D: T282-R238 G-protein coupled receptors signature: F102-I147 signature: T109-I125	BLAST- BLIMPS- PRINTS
15	747214ICD1	309	S266 T87 T162 T291	N4		HMMER HMMER-PP4M BLIMPS- BLOCKS PROFILESCAN MOTIFS

WO 01/98351

PCT/US01/19275

Table 3 (cont.)

SEQ ID NO: 16	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17	7477934CD1	326	S186 S20 S235 S289 S36 S65 T190 W191	N153 N5	OLFACTORY RECEPTOR PROTEIN PD000921: L164-L244 G-PROTEIN COUPLED RECEPTORS PD00013 F0995 18-305: L24-L24 7 transmembrane receptor (rhodopsin family) Tm.1: E39-L288 G-PROTEIN COUPLED RECEPTORS PD00013 A57069 15-304: F33-L303 RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: F166-L244 OLFACTORY RECEPTOR PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PDI49621: V245-F307 G-protein coupled receptor BL00237: R88-F127, L205-Y216, T280-L296 Olfactory receptor: signature PRO0245: F236-G251, L472-L283, S289-L303, M57-K78, F175-D189 G-protein coupled receptors signature: F100-G145 transmembrane domain: L25-L46 G_Protein_Receptor: GI08-I124	BLAST-PRODOM BLAST-DMO HMMER-FFAM BLAST-DMO BLAST-PRODOM BLAST-PRODOM BLTMP8-BLOCKS BLIMPS-PRINTS PROFILINSCAN HMMER MOTIFS

WO 01/98351

PCT/US01/19275

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
18	1714538CB1	2374	1-1652,	71687857V1	2141	2287
			2139-2374,	71688177V1	986	1547
			1-234,	71689955V1	1246	1643
19	3406743CB1	813	818-1572,	1714538H1 (UCMCMOT02)	4	289
			532-1151,	GNN:96855250_000025_002_edit	87	1365
			68-691	4516658H1 (STJNMT03)	2113	2374
			1-168,	FL210313_00001	1	813
			321-725			
20	3485895CB1	1542	1-334,	71296333V1	926	1542
			626-883,	71755255V1	525	1186
			1518-1542	60149346D1	87	518
				GNN:96479070_edit	1	300
21	7476102CB1	1191	1168-1191,	71295925V1	492	1180
			1-1097	GBI:97684460_edit	1	1191
22	2432942CB1	3360	1-1940,	2432516H1 (BRAVINT02)	1	238
			3316-3360,	70572441V1	1447	2041
			2803-2843	70254651V1	997	1498
				8116752H1 (TOKSDIC01)	682	1314
				7173326V1	300	966
				7183828V1	2675	3360
				55046035J2	331	943
				6544942F6 (LNDNRN02)	2018	2827
				71730213V1	2145	2854
				70815928V1	1502	2139
23	4630911CB1	1660	1-1260	2294242E6 (SPLNFF02)	1263	1660
				71047664V1	924	1516
				7705542V1 (UTRPFUE01)	245	563
				GBI:98112480_edit	1	1125
			6589031F8 (TLXAMNT03)	459	1158	

WO 01/98351

PCT/US01/19275

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
24	7472432CB1	2603	1-24,	6975661HH (BRANDR04)	1657	2249
			2540-2603,	5503722572	1	393
			838-2005	7341002HI (COLANDR02)	2336	2603
25	7474977CB1	1119	1-1119	7055089HI (BRALNR02)	1819	2477
			1-140,	FL7472432CB1_00001	239	1648
			262-344,	2564391F6 (SCORNR04)	1353	1728
			546-571,	97137574_edit	1	1119
26	7474848CB1	1018	1-140,	GB1_98081257_edit	1	1018
			262-344,			
			546-571,			
			992-1018			
35	7655614CB1.comp	2177	1537-2177,	7655614HI (UTREDR06)	653	1240
			1-39,	FL7655614_97406476_g1.56725	705	2177
			857-1072,	7655614JI (UTREDR06)	1	710
			372-472			
28	6792419CB1	1632	1-234,	6314357T6 (NERDR03)	811	1620
			1-695,	6110713F8 (RCLDXR03)	161	968
			752-809	94580923_CD	1	1632
29	7474790CB1	1458	1-103,	70619934V1	1048	1624
			370-1458	GNN_99759183_000017_002_edit	1	1458
30	7474816CB1	1015	1-489	55078072J1	1	596
			564-1015	FL7474816_98080066_000028_g1.055254	248	1015
31	7476172CB1	2781	888-2781,	FL7476172_g7684554_000005_95803686	1	2781
			1-252,			
			1-393			
32	7472141CB1	1267	582-2781			
			1-449,	97940222_edit	1	1267
33	7472137CB1	1559	1-938			
			1268-1559,	FL7472137_g8118970_000006_g1.256393	624	1559
34	7477934CB1	981	1-1043	55033677HI (FLRAX017)	1	877
			83-92,	GB1_98568403_000012_edit	1	981
			909-981			

WO 01/98351

PCT/US01/19275

Table 5

Polynucleotide SEQ. ID. NO.	Incyte Project ID	Representative Library
18	171453CB1	BRSTN002
19	3406743CB1	PROST008
20	3485895CB1	KIDN0020
22	2432842CB1	BRAV002
23	4630911CB1	SPLAF002
24	7472432CB1	SCOR004
27	7655614CB1	PITUL001
28	6792419CB1	NERD003

WO 01/98351

PCT/US01/19275

Table 6

Library	Vector	Library Description
BRAVUNT02	PSFORT1	Library was constructed using polyA RNA isolated from separate populations of unstimulated astrocytes. The RNA was pooled for polyA RNA isolation and library construction.
BRSTW002	PIWCY	This normalized breast tissue library was constructed from 6.2 million independent clones from a pool of two libraries from two different donors. Starting RNA was made from breast tissue removed from a 46-year-old Caucasian female during a bilateral reduction mastoplasty (donor A), and from breast tissue removed from a 60-year-old Caucasian female during a bilateral reduction mastoplasty (donor B). Pathology indicated normal breast parenchyma, bilaterally (A) and bilateral mammary hypertrophy (B). Patient history included hypertrophy of breast, obesity, lumbago, and glaucoma (A) and joint pain in the shoulder, thyroid cyst, colon cancer, normal delivery and cervical cancer (B). Family history included cataract, osteoarthritis, uterine cancer, benign hypertension, hyperlipidemia, and alcoholic cirrhosis of the liver, cerebrovascular disease, and type II diabetes (A) and cerebrovascular accident, atherosclerotic coronary artery disease, colon cancer, type II diabetes, hyperlipidemia, depressive disorder, and Alzheimer's Disease. The library was normalized in two rounds using conditions adapted from Soares et al. (1994) PNAS 91:9228-9232 and Bonaldo et al. (1996) Genome Research 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
KIDNWT02	PIWCY	Library was constructed using RNA isolated from left kidney tissue removed from a 43-year-old Caucasian male during nephroureterectomy, regional lymph node excision, and unilateral left adrenalectomy. Pathology for the associated tumor tissue indicated a grade 2 renal cell carcinoma. Family history included atherosclerotic coronary artery disease.
PIVUDIR01	PCIM2.1	This random primed library was constructed using RNA isolated from pituitary gland tissue removed from a 70-year-old female who died from metastatic adenocarcinoma. Pathology for the brain indicated moderate Alzheimer disease and mild carotid and cerebral atherosclerosis. The cerebral hemispheres, frontal and temporal lobes, white matter, and hippocampus showed mild atrophy, bilaterally. There were numerous neurofibrillary tangles, neuritic and diffuse amyloid plaques deposited throughout most neocortical areas.
SPANFET02	PIWCY	Library was constructed using RNA isolated from spleen tissue removed from a Caucasian male fetus, who died at 23 weeks' gestation.

WO 01/98351

PCT/US01/19275

Table 6 (cont.)

Library	Vector	Library Description
NERDUDM03	P1N1CY	This normalized dorsal root ganglion tissue library was constructed from 1.05 million independent clones from a dorsal root ganglion tissue library. Starting RNA was made from dorsal root ganglion tissue removed from the cervical spine of a 32-year-old Caucasian male who died from acute pulmonary edema, acute bronchopneumonia, bilateral pleural effusions, pericardial effusion, and malignant lymphoma (natural killer cell type). The patient presented with pyrexia of unknown origin, malaise, fatigue, and gastrointestinal bleeding. Patient history included probable cytomegalovirus infection, liver congestion, failure, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, respiratory failure, pneumonia of the left lung, natural killer cell lymphoma of the pharynx, Bell's palsy, and tobacco and alcohol abuse. Previous surgeries included colonoscopy, cecocolon biopsy, adenotomyllectomy, and nasopharyngeal endoscopy and biopsy. Patient medications included Diflucan (fluconazole), Peltason (prednisone), hydrocodone, Loratab, Alprazolam, Resazodone, ProMace-Cytabom, Etoposide, Cisplatin, Cytarabine, and dexamethasone. The patient received radiation therapy and multiple blood transfusions. The library was normalized in 2 rounds using conditions adapted from Soares et al. (1994) PNAS 91:9228-9232 and Bonaldo et al. (1996) Genome Research 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
PROSTUS08	PTT73	This subtracted library was constructed using 2.36 million clones from a prostate tumor library and was subjected to one round of subtractive hybridization with 448,000 clones from a normal prostate library. The starting library for subtraction was constructed using RNA isolated from a prostate tumor removed from a 59-year-old Caucasian male during a radical prostatectomy with regional lymph node excision. Pathology indicated adenocarcinoma (Gleason grade 3+3) Adenofibromatous hyperplasia was present. The patient presented with elevated prostate-specific antigen (PSA). Patient history included colon diverticuli, asbestosis, and thrombophlebitis. Family history included multiple myeloma, hyperlipidemia, and rheumatoid arthritis. Subtractive hybridization conditions were based on the methodologies of Swarcop et al. (1991) NAR 19:1954 and Bonaldo et al. (1996) Genome Research 6:791.
SCORNT04	P1N1CY	Library was constructed using RNA isolated from cervical spinal cord tissue removed from a 32-year-old Caucasian male who died from acute pulmonary edema and bronchopneumonia, bilateral pleural and pericardial effusions, and malignant lymphoma (natural killer cell type). Patient history included probable cytomegalovirus infection, hepatic congestion and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, and Bell's palsy. Surgeries included colonoscopy, large intestine biopsy, adenotomyllectomy, and nasopharyngeal endoscopy and biopsy. Treatment included radiation therapy.

WO 01/98351

PCT/US01/19275

Table 7

Program	Description	Reference	Parameter Threshold
ABI/FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL/DF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <5%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) <i>J. Mol. Biol.</i> 215:403-410; Altschul, S.F. et al. (1997) <i>Nucleic Acids Res.</i> 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, wfasta, tfastx, and tsearch.	Pearson, W.R. and D.J. Lipman (1988) <i>Proc. Natl. Acad. Sci. USA</i> 85:2444-2448; Pearson, W.R. (1990) <i>Methods Enzymol.</i> 183:63-98; and Smith, T.F. and M.S. Waterman (1981) <i>Adv. Appl. Math.</i> 2:482-489.	ESTs: fasta E value=1.0E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fasta E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) <i>Nucleic Acids Res.</i> 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) <i>Methods Enzymol.</i> 266:88-105; and Atwood, T.K. et al. (1997) <i>J. Chem. Inf. Comput. Sci.</i> 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) <i>J. Mol. Biol.</i> 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) <i>Nucleic Acids Res.</i> 26:370-372; Durbin, R. et al. (1998) <i>Our World View</i> , in a Nusselt, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less Signal peptide hits: Score= 0 or greater

100

WO 01/98351

PCT/US01/19275

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfilesScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality scores:GGC-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phix 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.	Scores= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SFSscan	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; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, 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 M51-59, Genetics Computer Group, Madison, WI.	

WO 01/98351

PCT/US01/19275

Table 8

Tissues	Polynucleotide SEQ ID NO:
Breast, Fat, Skin	33
Muscle, Bone, Synovium,	-
Connective tissue	+
Pancreas, Liver, Gallbladder	-
Brain: Amygdala, Thalamus, Hippocampus,	-
Entorhinal cortex, Archacocortex	-
Brain: Striatum, Caudate nucleus,	-
Putamen, Dentate nucleus,	-
Globus pallidus, Substantia innominata,	-
Ralphie magnus	-
Kidney, Fetal colon, Small intestine,	-
Ileum, Esophagus	-
Fetal heart, Aorta, Coronary artery	-
Fetal lung, Adult lung	-
Placenta, Prostate, Uterus	-
Olfactory bulb	-

WO 01/98351

PCT/US01/19275

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
 - a) a polypeptide comprising an amino acid sequence selected from the group consisting of
5 SEQ ID NO:1-17,
 - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical
to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17,
 - c) a biologically active fragment of a polypeptide having an amino acid sequence selected
from the group consisting of SEQ ID NO:1-17, and
 - 10 d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from
the group consisting of SEQ ID NO:1-17.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-
17.
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:18-34.
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

WO 01/98351

PCT/US01/19275

- b) recovering the polypeptide so expressed.
10. An isolated antibody which specifically binds to a polypeptide of claim 1.
- 5 11. An isolated polynucleotide selected from the group consisting of:
- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34,
- b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34,
- 10 c) a polynucleotide complementary to a polynucleotide of a),
- d) a polynucleotide complementary to a polynucleotide of b), and
- e) an RNA equivalent of a)-d).
12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a
- 15 polynucleotide of claim 11.
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
- 20 comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
- b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.
- 25 14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.
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:
- 30 a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

WO 01/98351

PCT/US01/19275

16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

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

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:

- 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.

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:

- 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.

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.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

WO 01/98351

PCT/US01/19275

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

5

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

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissible for the activity of the polypeptide of claim 1,
10 b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a
15 compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- 20 a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
b) detecting altered expression of the target polynucleotide, and
c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

25

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

- a) treating a biological sample containing nucleic acids with the test compound;
b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific
30 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
d) comparing the amount of hybridization complex in the treated biological sample with the
35 amount of hybridization complex in an untreated biological sample, wherein a difference in the

WO 01/98351

PCT/US01/19275

amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

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

- a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

10

30. The antibody of claim 10, wherein the antibody is:

- a) a chimeric antibody,
- b) a single chain antibody,
- c) a Fab fragment,
- d) a F(ab')₂ fragment, or
- e) a humanized antibody.

15

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

20

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

25

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

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

30

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

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

WO 01/98351

PCT/US01/19275

b) isolating antibodies from said animal; and
c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17.

5

36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

10 38. A method of making a monoclonal antibody with the specificity of the antibody of claim 10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

15

b) isolating antibody producing cells from the animal;

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;

d) culturing the hybridoma cells; and

20

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17.

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

25

41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

30

43. A method for detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17 in a sample, comprising the steps of:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific

WO 01/98351

PCT/US01/19275

binding of the antibody and the polypeptide; and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17 in the sample.

5

44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17 from a sample, the method comprising:

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

10

b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17.

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

15

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

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

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

20

49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

25

51. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

52. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

53. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

30

54. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

55. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

WO 01/98351

PCT/US01/19275

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.
57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.
58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.
59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.
60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.
61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.
62. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:18.
63. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:19.
64. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:20.
65. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:21.
66. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:22.
67. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:23.
68. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:24.
69. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID

WO 01/98351

PCT/US01/19275

NO:25.

70. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:26.

5

71. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:27.

72. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
10 NO:28.

73. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:29.

15 74. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:30.

75. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:31.

20

76. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:32.

25 77. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:33.

78. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:34.

WO 01/98351

PCT/US01/19275

<110> INCYTE GENOMICS, INC.
 LAL, Preeti
 BAUGHN, Mariah R.
 HAFALIA, April J. A.
 NGUYEN, Danniell B.
 GANDHI, Ameena R.
 KALLICK, Deborah A.
 GRIFFIN, Jennifer A.
 YUE, Henry
 KERN, Farrah A.
 PATTERSON, Chandra
 LU, Dyung Aina M.
 TRIBOULEY, Catherine M.
 LU, Yan
 WALIA, Narinder K.
 GRAUL, Richard
 YAO, Monique G.
 YANG, Junming
 RAMKUMAR, Jayalaxmi
 AU-YOUNG, Janice
 ELLIOTT, Vicki S.
 HERNANDEZ, Roberto
 WALSH, Roderick T.
 BORONSKY, Mark L.
 THORNTON, Michael
 HE, Ann

<120> G-PROTEIN COUPLED RECEPTORS

<130> PI-0131 PCT

<140> To Be Assigned
 <141> Herewith

<150> 60/212,483; 60/213,954; 60/215,209; 60/216,595; 60/218,936;
 60/219,154; 60/220,141
 <151> 2000-06-16; 2000-06-22; 2000-06-29; 2000-07-07; 2000-07-14;
 2000-07-19; 2000-07-21

<160> 35
 <170> PERL Program

<210> 1
 <211> 426
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1714538CD1

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

WO 01/98351

PCT/US01/19275

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

<210> 2
<211> 259
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3406743CD1

<400> 2
Met Glu Asp Leu Phe Ser Pro Ser Ile Leu Pro Pro Ala Pro Asn
1 5 10 15
Ile Ser Val Pro Ile Leu Leu Gly Trp Gly Leu Asn Leu Thr Leu
20 25 30

WO 01/98351

PCT/US01/19275

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

<210> 3
 <211> 379
 <212> PFT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3485895CD1

<400> 3
 Met Gln Ile Gly His Leu Ala Gln His Trp Val Arg Ser Tyr Ile
 1 5 10 15
 Gln Leu Leu Ala Glu Phe Leu Ser Arg Asp Gln Val Phe Gln Gln
 20 25 30
 Asn Gly Tyr Gly Leu Thr Gln Gln Asn Leu Leu Asn Asn Tyr Asp
 35 40 45
 Met Leu Gly Ile Met Ala Trp Asn Ala Thr Cys Lys Asn Trp Leu
 50 55 60
 Ala Ala Glu Ala Ala Leu Glu Lys Tyr Tyr Leu Ser Ile Phe Tyr
 65 70 75
 Gly Ile Glu Phe Val Val Gly Val Leu Gly Asn Thr Ile Val Val
 80 85 90
 Tyr Gly Tyr Ile Phe Ser Leu Lys Asn Trp Asn Ser Ser Asn Ile
 95 100 105
 Tyr Leu Phe Asn Leu Ser Val Ser Asp Leu Ala Phe Leu Cys Thr
 110 115 120
 Leu Pro Met Leu Ile Arg Ser Tyr Ala Asn Gly Asn Trp Ile Tyr
 125 130 135
 Gly Asp Val Leu Cys Ile Ser Asn Arg Tyr Val Leu His Ala Asn
 140 145 150

WO 01/98351

PCT/US01/19275

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

<210> 4

<211> 396

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7476102CD1

<400> 4

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

WO 01/98351

PCT/US01/19275

Ile His Pro Leu Arg Tyr Leu Ser Phe Met Ser His Gly Ala Ala
 155 160
 Trp Lys Ala Val Ala Leu Ile Trp Leu Val Ala Cys Cys Phe Pro
 170 175
 Thr Phe Leu Ile Trp Leu Ser Lys Trp Gln Asp Ala Gln Leu Glu
 185 190
 Glu Gln Gly Ala Ser Tyr Ile Leu Pro Pro Ser Met Gly Thr Gln
 200 205
 Pro Gly Cys Gly Leu Leu Val Ile Val Thr Tyr Thr Ser Ile Leu
 215 220
 Cys Val Leu Phe Leu Cys Thr Ala Leu Ile Ala Asn Cys Phe Trp
 230 235
 Arg Ile Tyr Ala Glu Ala Lys Thr Ser Gly Ile Trp Gly Gln Gly
 245 250
 Tyr Ser Arg Ala Arg Gly Thr Leu Leu Ile His Ser Val Leu Ile
 260 265
 Thr Leu Tyr Val Ser Thr Gly Val Val Phe Ser Leu Asp Met Val
 275 280
 Leu Thr Arg Tyr His His Ile Asp Ser Gly Thr His Thr Trp Leu
 290 295
 Leu Ala Ala Asn Ser Glu Val Leu Met Met Leu Pro Arg Ala Met
 305 310
 Leu Thr Tyr Leu Tyr Leu Leu Arg Tyr Arg Gln Leu Leu Gly Met
 320 325
 Val Arg Gly His Leu Pro Ser Arg Arg His Gln Ala Ile Phe Thr
 335 340
 Ile Ser Cys Cys Trp Glu Pro Leu Phe Tyr Phe Ser Leu Ile Leu
 350 355
 Thr Thr Leu Gly Val Asp Ile Ile Pro Leu Cys Val Glu Thr Thr
 365 370
 Phe Cys Leu Ser Ile His Leu Ser Met Asp Arg Leu Gly Cys Thr
 380 385
 Thr Phe Gly Tyr Cys Glu
 395

<210> 5

<211> 528

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2432942CD1

<400> 5

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

WO 01/98351

PCT/US01/19275

Phe Ser Asn Thr His Phe Phe Lys Asp Glu Asn Asn Ser Ser Leu
 140 145 150
 Leu Asn Asn Tyr Val Leu Gly Ala Gln Leu Ser His Gly His Val
 155 160 165
 Asn Asn Leu Arg Asp Pro Val Asn Ile Ser Phe Trp His Asn Gln
 170 175 180
 Ser Leu Glu Gly Tyr Thr Leu Thr Cys Val Phe Trp Lys Glu Gly
 185 190 195
 Ala Arg Lys Gln Pro Trp Gly Gly Trp Ser Pro Glu Gly Cys Arg
 200 205 210
 Thr Glu Gln Pro Ser His Ser Gln Val Leu Cys Arg Cys Asn His
 215 220 225
 Leu Thr Tyr Phe Ala Val Leu Met Gln Leu Ser Pro Ala Leu Val
 230 235 240
 Pro Ala Glu Leu Leu Ala Pro Leu Thr Tyr Ile Ser Leu Val Gly
 245 250 255
 Cys Ser Ile Ser Ile Val Ala Ser Leu Ile Thr Val Leu Leu His
 260 265 270
 Phe His Phe Arg Lys Gln Ser Asp Ser Leu Thr Arg Ile His Met
 275 280 285
 Asn Leu His Ala Ser Val Leu Leu Leu Asn Ile Ala Phe Leu Leu
 290 295 300
 Ser Pro Ala Phe Ala Met Ser Pro Val Pro Gly Ser Ala Cys Thr
 305 310 315
 Ala Leu Ala Ala Ala Leu His Tyr Ala Leu Leu Ser Cys Leu Thr
 320 325 330
 Trp Met Ala Ile Glu Gly Phe Asn Leu Tyr Leu Leu Leu Gly Arg
 335 340 345
 Val Tyr Asn Ile Tyr Ile Arg Arg Tyr Val Phe Lys Leu Gly Val
 350 355 360
 Leu Gly Trp Gly Ala Pro Ala Leu Leu Val Leu Leu Ser Leu Ser
 365 370 375
 Val Lys Ser Ser Val Tyr Gly Pro Cys Thr Ile Pro Val Phe Asp
 380 385 390
 Ser Trp Glu Asn Gly Thr Gly Phe Gln Asn Met Ser Ile Cys Trp
 395 400 405
 Val Arg Ser Pro Val Val His Ser Val Leu Val Met Gly Tyr Gly
 410 415 420
 Gly Leu Thr Ser Leu Phe Asn Leu Val Val Leu Ala Trp Ala Leu
 425 430 435
 Trp Thr Leu Arg Arg Leu Arg Glu Arg Ala Asp Ala Pro Ser Val
 440 445 450
 Arg Ala Cys His Asp Thr Val Thr Val Leu Gly Leu Thr Val Leu
 455 460 465
 Leu Gly Thr Thr Trp Ala Leu Ala Phe Phe Ser Phe Gly Val Phe
 470 475 480
 Leu Leu Pro Gln Leu Phe Leu Phe Thr Ile Leu Asn Ser Leu Tyr
 485 490 495
 Gly Phe Phe Leu Phe Leu Trp Phe Cys Ser Gln Arg Cys Arg Ser
 500 505 510
 Glu Ala Glu Ala Lys Ala Gln Ile Glu Ala Phe Ser Ser Ser Gln
 515 520 525
 Thr Thr Gln

<210> 6

<211> 361

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 4630911CD1

WO 01/98351

PCT/US01/19275

```

<400> 6
Met Asn Arg Lys Gln Leu Glu Val Ile Pro Met Ile Trp Val Ile
1 5 10 15
Ser Leu Asp Lys Ser Lys Gly Thr Phe His Ile Gly Ser His Phe
20 25 30
Thr Ser Thr Ser Leu Ile Phe Ser Asn Cys Thr Asn Thr Asp Phe
35 40 45
Arg Tyr Phe Ile Tyr Ala Val Thr Tyr Thr Val Ile Leu Val Pro
50 55 60
Gly Leu Ile Gly Asn Ile Leu Ala Leu Trp Val Phe Tyr Gly Tyr
65 70 75
Met Lys Glu Thr Lys Arg Ala Val Ile Phe Met Ile Asn Leu Ala
80 85 90
Ile Ala Asp Leu Leu Gln Val Leu Ser Leu Pro Leu Arg Ile Phe
95 100 105
Tyr Tyr Leu Asn His Asp Trp Pro Phe Gly Pro Gly Leu Cys Met
110 115 120
Phe Cys Phe Tyr Leu Lys Tyr Val Asn Met Tyr Ala Ser Ile Tyr
125 130 135
Phe Leu Val Cys Ile Ser Val Arg Arg Phe Trp Phe Leu Met Tyr
140 145 150
Pro Phe Arg Phe His Asp Cys Lys Gln Lys Tyr Asp Leu Tyr Ile
155 160 165
Ser Ile Ala Gly Trp Leu Ile Ile Cys Leu Ala Cys Val Leu Phe
170 175 180
Pro Leu Leu Arg Thr Ser Asp Asp Thr Pro Gly Asn Arg Thr Lys
185 190 195
Cys Phe Val Asp Leu Pro Thr Arg Asn Val Asn Leu Ala Gln Ser
200 205 210
Val Val Met Met Thr Ile Gly Glu Leu Ile Gly Phe Val Thr Pro
215 220 225
Leu Leu Ile Val Leu Tyr Cys Thr Trp Lys Thr Val Leu Ser Leu
230 235 240
Gln Asp Lys Tyr Pro Met Ala Gln Asp Leu Gly Glu Lys Gln Lys
245 250 255
Ala Leu Lys Met Ile Leu Thr Cys Ala Gly Val Phe Leu Ile Cys
260 265 270
Phe Ala Pro Tyr His Phe Ser Phe Pro Leu Asp Phe Leu Val Lys
275 280 285
Ser Asn Glu Ile Lys Ser Cys Leu Ala Arg Arg Val Ile Leu Ile
290 295 300
Phe His Ser Val Ala Leu Cys Leu Ala Ser Leu Asn Ser Cys Leu
305 310 315
Asp Pro Val Ile Tyr Tyr Phe Ser Thr Asn Glu Phe Arg Arg Arg
320 325 330
Leu Ser Arg Gln Asp Leu His Asp Ser Ile Gln Leu His Ala Lys
335 340 345
Ser Phe Val Ser Asn His Thr Ala Ser Thr Met Thr Pro Glu Leu
350 355 360
Cys

```

```

<210> 7
<211> 469
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472432CD1

```

```

<400> 7
Met Ala Phe Leu Met His Leu Leu Val Cys Val Phe Gly Met Gly

```

WO 01/98351

PCT/US01/19275

```

1           5           10           15
Ser Trp Val Thr Ile Asn Gly Leu Trp Val Glu Leu Pro Leu Leu
20
Val Met Glu Leu Pro Glu Gly Trp Tyr Leu Pro Ser Tyr Leu Thr
35
Val Val Ile Gln Leu Ala Asn Ile Gly Pro Leu Leu Val Thr Leu
50
Leu His His Phe Arg Pro Ser Cys Leu Ser Glu Val Pro Met Ile
65
Phe Thr Leu Leu Gly Val Gly Thr Val Thr Cys Ile Ile Phe Ala
80
Phe Leu Trp Asn Met Thr Ser Trp Val Leu Asp Gly His His Ser
95
Ile Ala Phe Leu Val Leu Thr Phe Phe Leu Ala Leu Val Asp Cys
110
Thr Ser Ser Val Thr Phe Leu Pro Phe Met Ser Arg Leu Pro Thr
125
Tyr Tyr Leu Thr Thr Phe Phe Val Gly Glu Gly Leu Ser Gly Leu
140
Leu Pro Ala Leu Val Ala Leu Ala Gln Gly Ser Gly Leu Thr Thr
155
Cys Val Asn Val Thr Glu Ile Ser Asp Ser Val Pro Ser Pro Val
170
Pro Thr Arg Glu Thr Asp Ile Ala Gln Gly Val Pro Arg Ala Leu
185
Val Ser Ala Leu Pro Gly Met Glu Ala Pro Leu Ser His Leu Glu
200
Ser Arg Tyr Leu Pro Ala His Phe Ser Pro Leu Val Phe Phe Leu
215
Leu Leu Ser Ile Met Ala Cys Cys Leu Val Ala Phe Phe Val
230
Leu Gln Arg Gln Pro Arg Cys Trp Glu Ala Ser Val Glu Asp Leu
245
Leu Asn Asp Gln Val Thr Leu His Ser Ile Arg Pro Arg Glu Glu
260
Asn Asp Leu Gly Pro Ala Gly Thr Val Asp Ser Ser Gln Gly Gln
275
Gly Tyr Leu Glu Glu Iys Ala Ala Pro Cys Cys Pro Ala His Leu
290
Ala Phe Ile Tyr Thr Leu Val Ala Phe Val Asn Ala Leu Thr Asn
305
Gly Met Leu Pro Ser Val Gln Thr Tyr Ser Cys Leu Ser Tyr Gly
320
Pro Val Ala Tyr His Leu Ala Ala Thr Leu Ser Ile Val Ala Asn
335
Pro Leu Ala Ser Leu Val Ser Met Phe Leu Pro Asn Arg Ser Leu
350
Leu Phe Leu Gly Val Leu Ser Val Leu Gly Thr Cys Phe Gly Gly
365
Tyr Asn Met Ala Met Ala Val Met Ser Pro Cys Pro Leu Leu Gln
380
Gly His Trp Gly Gly Glu Val Leu Ile Val Ala Ser Trp Val Leu
395
Phe Ser Gly Cys Leu Ser Tyr Val Lys Val Met Leu Gly Val Val
410
Leu Arg Asp Leu Ser Arg Ser Ala Leu Leu Trp Cys Gly Ala Ala
425
Val Gln Leu Gly Ser Leu Leu Gly Ala Leu Leu Met Phe Pro Leu
440
Val Asn Val Leu Arg Leu Phe Ser Ser Ala Asp Phe Cys Asn Leu
455
His Cys Pro Ala
460

```

WO 01/98351

PCT/US01/19275

<210> 8
<211> 372
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474977CD1

```

<400> 8
Met Glu Ala Ala Ser Leu Ser Val Ala Thr Ala Gly Val Ala Leu
1 5 10 15
Ala Leu Gly Pro Glu Thr Ser Ser Gly Thr Pro Ser Pro Arg Gly
20 25 30
Ile Leu Gly Ser Thr Pro Ser Gly Ala Val Leu Pro Gly Arg Gly
35 40 45
Pro Pro Phe Ser Val Phe Thr Val Leu Val Val Thr Leu Leu Val
50 55 60
Leu Leu Ile Ala Ala Thr Phe Leu Trp Asn Leu Leu Val Pro Val
65 70 75
Thr Ile Pro Arg Val Arg Ala Phe His Arg Val Pro His Asn Leu
80 85 90
Val Ala Ser Thr Ala Val Ser Asp Glu Leu Val Ala Ala Leu Ala
95 100 105
Met Pro Pro Ser Leu Ala Ser Glu Leu Ser Thr Gly Arg Arg Arg
110 115 120
Leu Leu Gly Arg Ser Leu Cys His Val Trp Ile Ser Phe His Val
125 130 135
Leu Cys Cys Pro Ala Gly Leu Gly Asn Val Ala Ala Ile Ala Leu
140 145 150
Gly Arg Asp Gly Ala Ile Thr Arg His Leu Gln His Thr Leu Arg
155 160 165
Thr Arg Ser Arg Ala Ser Leu Leu Met Ile Ala Leu Thr Arg Val
170 175 180
Pro Ser Ala Leu Ile Ala Leu Ala Pro Leu Leu Phe Gly Arg Gly
185 190 195
Glu Val Cys Asp Ala Arg Leu Gln Arg Cys Gln Val Ser Arg Glu
200 205 210
Pro Ser Tyr Ala Ala Phe Ser Thr Arg Gly Ala Phe His Leu Pro
215 220 225
Leu Gly Val Val Pro Phe Val Tyr Arg Lys Ile Tyr Glu Ala Ala
230 235 240
Lys Phe Arg Phe Gly Arg Arg Arg Ala Val Leu Pro Leu Pro
245 250 255
Ala Thr Met Gln Val Lys Val Lys Glu Ala Pro Asp Glu Ala Glu
260 265 270
Val Val Phe Thr Ala His Cys Lys Ala Thr Val Ser Phe Gln Val
275 280 285
Ser Gly Asp Ser Trp Arg Glu Gln Lys Glu Arg Arg Ala Ala Met
290 295 300
Met Val Gly Ile Leu Ile Gly Val Phe Val Leu Cys Trp Ile Pro
305 310 315
Phe Phe Leu Thr Glu Leu Ile Ser Pro Leu Cys Ala Cys Ser Leu
320 325 330
Pro Pro Ile Trp Lys Ser Ile Phe Leu Trp Leu Gly Tyr Ser Asn
335 340 345
Ser Phe Phe Asn Pro Leu Ile Tyr Thr Ala Phe Asn Lys Asn Tyr
350 355 360
Asn Asn Ala Phe Lys Ser Leu Phe Thr Lys Gln Arg
365 370

```

<210> 9
<211> 330

WO 01/98351

PCT/US01/19275

<212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7474848CD1

<400> 9
 Met Asp Pro Thr Thr Pro Ala Trp Gly Thr Glu Ser Thr Thr Val
 1 5 10 15
 Asn Gly Asn Asp Gln Ala Leu Leu Leu Leu Cys Gly Lys Glu Thr
 20 25 30
 Leu Ile Pro Val Phe Leu Ile Leu Phe Ile Ala Leu Val Gly Leu
 35 40 45
 Val Gly Asn Gly Phe Val Leu Trp Leu Leu Gly Phe Arg Met Arg
 50 55 60
 Arg Asn Ala Phe Ser Val Tyr Val Leu Ser Leu Ala Gly Ala Asp
 65 70 75
 Phe Leu Phe Leu Cys Phe Gln Ile Ile Asn Cys Leu Val Tyr Leu
 80 85 90
 Ser Asn Phe Phe Cys Ser Ile Ser Ile Asn Phe Pro Ser Phe Phe
 95 100 105
 Thr Thr Val Met Thr Cys Ala Tyr Leu Ala Gly Leu Ser Met Leu
 110 115 120
 Ser Thr Val Ser Thr Glu Arg Cys Leu Ser Val Leu Trp Pro Ile
 125 130 135
 Trp Tyr Arg Cys Arg Arg Pro Arg His Leu Ser Ala Val Val Cys
 140 145 150
 Val Leu Leu Trp Ala Leu Ser Leu Leu Ser Ile Leu Glu Gly
 155 160 165
 Lys Phe Cys Gly Phe Leu Phe Ser Asp Gly Asp Ser Gly Trp Cys
 170 175 180
 Gln Thr Phe Asp Phe Ile Thr Ala Ala Trp Leu Ile Phe Leu Phe
 185 190 195
 Met Val Leu Cys Gly Ser Ser Leu Ala Leu Leu Val Arg Ile Leu
 200 205 210
 Cys Gly Ser Arg Gly Leu Pro Leu Thr Arg Leu Tyr Leu Thr Ile
 215 220 225
 Leu Leu Thr Val Leu Val Phe Leu Leu Cys Gly Leu Pro Phe Gly
 230 235 240
 Ile Gln Trp Phe Leu Ile Leu Trp Ile Trp Lys Asp Ser Asp Val
 245 250 255
 Leu Phe Cys His Ile His Pro Val Ser Val Val Leu Ser Ser Leu
 260 265 270
 Asn Ser Ser Ala Asn Pro Ile Ile Tyr Phe Phe Val Gly Ser Phe
 275 280 285
 Arg Lys Gln Trp Arg Leu Gln Gln Pro Ile Leu Lys Leu Ala Leu
 290 295 300
 Gln Arg Ala Leu Gln Asp Ile Ala Glu Val Asp His Ser Glu Gly
 305 310 315
 Cys Phe Arg Gln Gly Thr Pro Glu Met Ser Arg Ser Ser Leu Val
 320 325 330

<210> 10
 <211> 494
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7655614CD1

WO 01/98351

PCT/US01/19275

```

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

```

11/31

WO 01/98351

PCT/US01/19275

Val	Val	Leu	Phe	Cys	Arg	Glu	Gly	Pro	Pro	Glu	Pro	Val	Thr	Ala
				470					475					480
Val	Thr	Lys	Gln	Pro	Lys	Ser	Glu	Ala	Gly	Asp	Thr	Ser	Leu	
				485					490					

<210> 11
 <211> 532
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 6792419CD1

<400> 11
 Met Asn Lys Ser Thr Cys Leu Met Ala Ala Glu Thr Pro Ser Lys
 1 5 10 15
 Arg Trp Arg Leu His Cys Leu Ala Phe Ser Gln Arg Phe Val Arg
 20 25 30
 Ala Gly Pro Ala Cys Ser Ser Arg Glu Ala Cys Ser Ser Pro Arg
 35 40 45
 Ala Gly Trp Asn Pro Ala Gly Phe Arg Leu Pro Gly Arg Trp Ser
 50 55 60
 Pro Phe Val Ala Leu His Leu Val Cys Gln Ile Arg Glu Ala Leu
 65 70 75
 Lys Leu Arg Ser Gly His Arg Thr Pro Ser Gly Ala Gly Ser Ser
 80 85 90
 Met Gln Pro Pro Pro Ser Leu Cys Gly Arg Ala Leu Val Ala Leu
 95 100 105
 Val Leu Ala Cys Gly Leu Ser Arg Ile Trp Gly Glu Glu Arg Gly
 110 115 120
 Phe Pro Pro Asp Arg Ala Thr Pro Leu Leu Gln Thr Ala Glu Ile
 125 130 135
 Met Thr Pro Pro Thr Lys Thr Leu Trp Pro Lys Gly Ser Asn Ala
 140 145 150
 Ser Leu Ala Arg Ser Leu Ala Pro Ala Glu Val Pro Lys Gly Asp
 155 160 165
 Arg Thr Ala Gly Ser Pro Pro Arg Thr Ile Ser Pro Pro Pro Cys
 170 175 180
 Gln Gly Pro Ile Glu Ile Lys Glu Thr Phe Lys Tyr Ile Asn Thr
 185 190 195
 Val Val Ser Cys Leu Val Phe Val Leu Gly Ile Ile Gly Asn Ser
 200 205 210
 Thr Leu Leu Arg Ile Ile Tyr Lys Asn Lys Cys Met Arg Asn Gly
 215 220 225
 Pro Asn Ile Leu Ile Ala Ser Leu Ala Leu Gly Asp Leu Leu His
 230 235 240
 Ile Val Ile Asp Ile Pro Ile Asn Val Tyr Lys Leu Leu Ala Glu
 245 250 255
 Asp Trp Pro Phe Gly Ala Glu Met Cys Lys Leu Val Pro Phe Ile
 260 265 270
 Gln Lys Ala Ser Val Gly Ile Thr Val Leu Ser Leu Cys Ala Leu
 275 280 285
 Ser Ile Asp Arg Tyr Arg Ala Val Ala Ser Trp Ser Arg Ile Lys
 290 295 300
 Gly Ile Gly Val Pro Lys Trp Thr Ala Val Glu Ile Val Leu Ile
 305 310 315
 Trp Val Val Ser Val Val Leu Ala Val Pro Glu Ala Ile Gly Phe
 320 325 330
 Asp Ile Ile Thr Met Asp Tyr Lys Gly Ser Tyr Leu Arg Ile Cys
 335 340 345
 Leu Leu His Pro Val Gln Lys Thr Ala Phe Met Gln Phe Tyr Lys
 350 355 360

WO 01/98351

PCT/US01/19275

Thr Ala Lys Asp Trp Trp Leu Phe Ser Phe Tyr Phe Cys Leu Pro
 365 370 375
 Leu Ala Ile Thr Ala Phe Phe Tyr Thr Leu Met Thr Cys Glu Met
 380 385 390
 Leu Arg Lys Lys Ser Gly Met Gln Ile Ala Leu Asn Asp His Leu
 395 400 405
 Lys Gln Arg Arg Glu Val Ala Lys Thr Val Phe Cys Leu Val Leu
 410 415 420
 Val Phe Ala Leu Cys Trp Leu Pro Leu His Leu Ser Arg Ile Leu
 425 430 435
 Lys Leu Thr Leu Tyr Asn Gln Asn Asp Pro Asn Arg Cys Glu Leu
 440 445 450
 Leu Ser Phe Leu Leu Val Leu Asp Tyr Ile Gly Ile Asn Met Ala
 455 460 465
 Ser Leu Asn Ser Cys Ile Asn Pro Ile Ala Leu Tyr Leu Val Ser
 470 475 480
 Lys Arg Phe Lys Asn Cys Phe Lys Ser Cys Leu Cys Cys Trp Cys
 485 490 495
 Gln Ser Phe Glu Glu Lys Gln Ser Leu Glu Glu Lys Gln Ser Cys
 500 505 510
 Leu Lys Phe Lys Ala Asn Asp His Gly Tyr Asp Asn Phe Arg Ser
 515 520 525
 Ser Asn Lys Tyr Ser Ser Ser
 530

<210> 12
 <211> 485
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7474790CD1

<400> 12
 Met Pro Ile Ser Leu Ala His Gly Ile Ile Arg Ser Thr Val Leu
 1 5 10 15
 Val Ile Phe Leu Ala Ala Ser Phe Val Gly Asn Ile Val Leu Ala
 20 25 30
 Leu Val Leu Gln Arg Lys Pro Gln Leu Leu Gln Val Thr Asn Arg
 35 40 45
 Phe Ile Phe Asn Leu Leu Val Thr Asp Leu Leu Gln Ile Ser Leu
 50 55 60
 Val Ala Pro Trp Val Val Ala Thr Ser Val Pro Leu Phe Trp Pro
 65 70 75
 Leu Asn Ser His Phe Cys Thr Ala Leu Val Ser Leu Thr His Leu
 80 85 90
 Phe Ala Phe Ala Ser Val Asn Thr Ile Val Val Val Ser Val Asp
 95 100 105
 Arg Tyr Leu Ser Ile Ile His Pro Leu Ser Tyr Pro Ser Lys Met
 110 115 120
 Thr Gln Arg Arg Gly Tyr Leu Leu Leu Tyr Gly Thr Trp Ile Val
 125 130 135
 Ala Ile Leu Gln Ser Thr Pro Pro Leu Tyr Gly Trp Gly Gln Ala
 140 145 150
 Ala Phe Asp Glu Arg Asn Ala Leu Cys Ser Met Ile Trp Gly Ala
 155 160 165
 Ser Pro Ser Tyr Thr Ile Leu Ser Val Val Ser Phe Ile Val Ile
 170 175 180
 Pro Leu Ile Val Met Ile Ala Cys Tyr Ser Val Val Phe Cys Ala
 185 190 195
 Ala Arg Arg Gln His Ala Leu Leu Tyr Asn Val Lys Arg His Ser
 200 205 210

WO 01/98351

PCT/US01/19275

Leu Glu Val Arg Val Lys Asp Cys Val Glu Asn Glu Asp Glu Glu
 215 220 225
 Gly Ala Glu Lys Lys Glu Glu Phe Gln Asp Glu Ser Glu Phe Arg
 230 235 240
 Arg Gln His Glu Gly Glu Val Lys Ala Lys Glu Gly Arg Met Glu
 245 250 255
 Ala Lys Asp Gly Ser Leu Lys Ala Lys Glu Gly Ser Thr Gly Thr
 260 265 270
 Ser Glu Ser Ser Val Glu Ala Arg Gly Ser Glu Glu Val Arg Glu
 275 280 285
 Ser Ser Thr Val Ala Ser Asp Gly Ser Met Glu Gly Lys Glu Gly
 290 295 300
 Ser Thr Lys Val Glu Glu Asn Ser Met Lys Ala Asp Lys Gly Arg
 305 310 315
 Thr Glu Val Asn Gln Cys Ser Ile Asp Leu Gly Glu Asp Asp Met
 320 325 330
 Glu Phe Gly Glu Asp Asp Ile Asn Phe Ser Glu Asp Asp Val Glu
 335 340 345
 Ala Val Asn Ile Pro Glu Ser Leu Pro Pro Ser Arg Arg Asn Ser
 350 355 360
 Asn Ser Asn Pro Pro Leu Pro Arg Cys Tyr Gln Cys Lys Ala Ala
 365 370 375
 Lys Val Ile Phe Ile Ile Ile Phe Ser Tyr Val Leu Ser Leu Gly
 380 385 390
 Pro Tyr Cys Phe Leu Ala Val Leu Ala Val Trp Val Asp Val Glu
 395 400 405
 Thr Gln Val Pro Gln Trp Val Ile Thr Ile Ile Ile Trp Leu Phe
 410 415 420
 Phe Leu Gln Cys Cys Ile His Pro Tyr Val Tyr Gly Tyr Met His
 425 430 435
 Lys Thr Ile Lys Lys Glu Ile Gln Asp Met Leu Lys Lys Phe Phe
 440 445 450
 Cys Lys Glu Lys Pro Pro Lys Glu Asp Ser His Pro Asp Leu Pro
 455 460 465
 Gly Thr Glu Gly Gly Thr Glu Gly Lys Ile Val Pro Ser Tyr Asp
 470 475 480
 Ser Ala Thr Phe Pro
 485

<210> 13
 <211> 255
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7474816CD1

<400> 13
 Met Pro Phe Ile Ser Lys Leu Val Leu Ala Ser Gln Pro Thr Leu
 1 5 10 15
 Phe Ser Phe Phe Ser Ala Ser Ser Pro Phe Leu Leu Phe Leu Asp
 20 25 30
 Leu Arg Pro Glu Arg Thr Tyr Leu Pro Val Cys His Val Ala Leu
 35 40 45
 Ile His Met Val Val Leu Leu Thr Met Val Phe Leu Ser Pro Gln
 50 55 60
 Leu Phe Glu Ser Leu Asn Phe Gln Asn Asp Phe Lys Tyr Glu Ala
 65 70 75
 Ser Phe Tyr Leu Arg Arg Val Ile Arg Asp Leu Ser Ile Cys Thr
 80 85 90
 Thr Cys Leu Leu Gly Met Leu Gln Val Val Asn Ile Ser Pro Ser
 95 100 105

WO 01/98351

PCT/US01/19275

```

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

```

```

<210> 14
<211> 881
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7476172CD1

```

```

<400> 14
Met Ser Ile Glu Glu Leu Cys Ser Asp Phe Lys Lys Tyr Leu Phe
1 5 10 15
Pro Asn Ser Phe Glu Ile Ser Val Phe Leu Gln Thr Leu Ala Met
20 25 30
Ile His Ser Ile Glu Met Ile Asn Asn Ser Thr Leu Leu Pro Gly
35 40 45
Val Lys Leu Gly Tyr Glu Ile Tyr Asp Thr Cys Thr Glu Val Thr
50 55 60
Val Ala Met Ala Ala Thr Leu Arg Phe Leu Ser Lys Phe Asn Cys
65 70 75
Ser Arg Glu Thr Val Glu Phe Lys Cys Asp Tyr Ser Ser Tyr Met
80 85 90
Pro Arg Val Lys Ala Val Ile Gly Ser Gly Tyr Ser Glu Ile Thr
95 100 105
Met Ala Val Ser Arg Met Leu Asn Leu Leu Met Pro Gln Val
110 115 120
Gly Tyr Glu Ser Thr Ala Glu Ile Leu Ser Asp Lys Ile Arg Phe
125 130 135
Pro Ser Phe Leu Arg Thr Val Pro Ser Asp Phe His Gln Ile Lys
140 145 150
Ala Met Ala His Leu Ile Gln Lys Ser Gly Trp Asn Trp Ile Gly
155 160 165
Ile Ile Thr Thr Asp Asp Asp Tyr Gly Arg Leu Ala Leu Asn Thr
170 175 180
Phe Ile Ile Gln Ala Glu Ala Asn Asn Val Cys Ile Ala Phe Lys
185 190 195
Glu Val Leu Pro Ala Phe Leu Ser Asp Asn Thr Ile Glu Val Arg
200 205 210
Ile Asn Arg Thr Leu Lys Lys Ile Ile Leu Glu Ala Gln Val Asn
215 220 225
Val Ile Val Val Phe Leu Arg Gln Phe His Val Phe Asp Leu Phe

```

WO 01/98351

PCT/US01/19275

	230		235		240									
Asn	Lys	Ala	Ile	Glu	Met	Asn	Ile	Asn	Lys	Met	Trp	Ile	Ala	Ser
	245		250		255									
Asp	Asn	Trp	Ser	Thr	Ala	Thr	Lys	Ile	Thr	Thr	Ile	Pro	Asn	Val
	260		265		270									
Lys	Lys	Ile	Gly	Lys	Val	Val	Gly	Phe	Ala	Phe	Arg	Arg	Gly	Asn
	275		280		285									
Ile	Ser	Ser	Phe	His	Ser	Phe	Leu	Gln	Asn	Leu	His	Leu	Leu	Pro
	290		295		300									
Ser	Asp	Ser	His	Lys	Leu	Leu	His	Glu	Tyr	Ala	Met	His	Leu	Ser
	305		310		315									
Ala	Cys	Ala	Tyr	Val	Lys	Asp	Thr	Asp	Leu	Ser	Gln	Cys	Ile	Phe
	320		325		330									
Asn	His	Ser	Gln	Arg	Thr	Leu	Ala	Tyr	Lys	Ala	Asn	Lys	Ala	Ile
	335		340		345									
Glu	Arg	Asn	Phe	Val	Met	Arg	Asn	Asp	Phe	Leu	Trp	Asp	Tyr	Ala
	350		355		360									
Glu	Pro	Gly	Leu	Ile	His	Ser	Ile	Gln	Leu	Ala	Val	Phe	Ala	Leu
	365		370		375									
Gly	Tyr	Ala	Ile	Arg	Asp	Leu	Cys	Gln	Ala	Arg	Asp	Cys	Gln	Asn
	380		385		390									
Pro	Asn	Ala	Phe	Gln	Pro	Trp	Glu	Leu	Leu	Gly	Val	Leu	Lys	Asn
	395		400		405									
Val	Thr	Phe	Thr	Asp	Gly	Trp	Asn	Ser	Phe	His	Phe	Asp	Ala	His
	410		415		420									
Gly	Asp	Leu	Asn	Thr	Gly	Tyr	Asp	Val	Val	Leu	Trp	Lys	Glu	Ile
	425		430		435									
Asn	Gly	His	Met	Thr	Val	Thr	Lys	Met	Ala	Glu	Tyr	Asp	Leu	Gln
	440		445		450									
Asn	Asp	Val	Phe	Ile	Ile	Pro	Asp	Gln	Glu	Thr	Lys	Asn	Glu	Phe
	455		460		465									
Arg	Asn	Leu	Lys	Leu	Thr	Leu	Phe	Ser	Val	Leu	Thr	Lys	Leu	Lys
	470		475		480									
His	Gln	Lys	Arg	Ile	Pro	Val	Ala	Thr	Val	Thr	Ser	Val	Pro	Val
	485		490		495									
Pro	Leu	Pro	Ser	Ile	Trp	His	Tyr	Arg	Gln	Thr	Val	Cys	Ala	Pro
	500		505		510									
Ser	Gln	Asp	Met	Pro	His	Cys	Leu	Leu	Cys	Asn	Asn	Lys	Thr	His
	515		520		525									
Trp	Ala	Pro	Val	Arg	Ser	Thr	Met	Cys	Phe	Glu	Lys	Glu	Val	Glu
	530		535		540									
Tyr	Leu	Asn	Trp	Asn	Asp	Ser	Leu	Ala	Ile	Leu	Leu	Leu	Ile	Leu
	545		550		555									
Ser	Leu	Leu	Gly	Ile	Ile	Phe	Val	Leu	Val	Val	Gly	Ile	Ile	Phe
	560		565		570									
Thr	Arg	Asn	Leu	Asn	Thr	Pro	Val	Val	Lys	Ser	Ser	Gly	Gly	Leu
	575		580		585									
Arg	Val	Cys	Tyr	Val	Ile	Leu	Leu	Cys	His	Phe	Leu	Asn	Phe	Ala
	590		595		600									
Ser	Thr	Ser	Phe	Phe	Ile	Gly	Glu	Pro	Gln	Asp	Phe	Thr	Cys	Lys
	605		610		615									
Thr	Arg	Gln	Thr	Met	Phe	Gly	Val	Ser	Phe	Thr	Leu	Cys	Ile	Ser
	620		625		630									
Cys	Ile	Leu	Thr	Lys	Ser	Leu	Lys	Ile	Leu	Leu	Ala	Phe	Ser	Phe
	635		640		645									
Asp	Pro	Lys	Leu	Gln	Lys	Phe	Leu	Lys	Cys	Leu	Tyr	Arg	Pro	Ile
	650		655		660									
Leu	Ile	Ile	Phe	Thr	Cys	Thr	Gly	Ile	Gln	Val	Val	Ile	Cys	Thr
	665		670		675									
Leu	Trp	Leu	Ile	Phe	Ala	Ala	Pro	Thr	Val	Glu	Val	Asn	Val	Ser
	680		685		690									
Leu	Pro	Arg	Val	Ile	Ile	Leu	Glu	Cys	Glu	Glu	Gly	Ser	Ile	Leu
	695		700		705									

WO 01/98351

PCT/US01/19275

Ala Phe Gly Thr Met Leu Gly Tyr Ile Ala Ile Leu Ala Phe Ile 710 715 720
 Cys Phe Ile Phe Ala Phe Lys Gly Lys Tyr Glu Asn Tyr Asn Glu 725 730 735
 Ala Lys Phe Ile Thr Phe Gly Met Leu Ile Tyr Phe Ile Ala Trp 740 745 750
 Ile Thr Phe Ile Pro Ile Tyr Ala Thr Thr Phe Gly Lys Tyr Val 755 760 765
 Pro Ala Val Glu Ile Ile Val Ile Leu Ile Ser Asn Tyr Gly Ile 770 775 780
 Leu Tyr Cys Thr Phe Ile Pro Lys Cys Tyr Val Ile Ile Cys Lys 785 790 795
 Gln Glu Ile Asn Thr Lys Ser Ala Phe Leu Lys Met Ile Tyr Ser 800 805 810
 Tyr Ser Ser His Ser Val Ser Ser Ile Ala Leu Ser Pro Ala Ser 815 820 825
 Leu Asp Ser Met Ser Gly Asn Val Thr Met Thr Asn Pro Ser Ser 830 835 840
 Ser Gly Lys Ser Ala Thr Trp Gln Lys Ser Lys Asp Leu Gln Ala 845 850 855
 Gln Ala Phe Ala His Ile Cys Arg Glu Asn Ala Thr Ser Val Ser 860 865 870
 Lys Thr Leu Pro Arg Lys Arg Met Ser Ser Ile 875 880

<210> 15
 <211> 309
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472141CD1

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

17/31

WO 01/98351

PCT/US01/19275

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

<210> 16
 <211> 224
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472137CD1

<400> 16
 Met Arg Asn Phe Ser Val Val Ser Glu Phe Ile Leu Leu Gly Ile 15
 1 5 10 15
 Pro His Thr Glu Gly Leu Glu Thr Ile Leu Val Leu Phe Leu 30
 20 25 30
 Ser Phe Tyr Ile Phe Thr Leu Met Gly Asn Leu Leu Ile Leu Leu 45
 35 40 45
 Ala Ile Val Ser Ser Ala Arg Leu His Thr Pro Met Tyr Phe Phe 60
 50 55 60
 Leu Cys Lys Leu Ser Val Phe Asp Leu Phe Phe Pro Ser Val Ser 75
 65 70 75
 Ser Pro Lys Met Leu Cys Tyr Leu Ser Gly Asn Ser Arg Ala Ile 90
 80 85 90
 Ser Tyr Ala Gly Cys Ala Ser Gln Leu Phe Phe Tyr His Phe Leu 105
 95 100 105
 Gly Cys Thr Glu Cys Phe Leu Tyr Thr Val Met Ala Tyr Asp Arg 120
 110 115 120
 Phe Val Ala Ile Cys His Pro Leu Arg Tyr Thr Ile Ile Met Ser 135
 125 130 135
 His Arg Ala Cys Ile Ile Leu Ala Met Gly Thr Ser Phe Phe Gly 150
 140 145 150
 Cys Ile Gln Ala Thr Phe Leu Thr Thr Leu Thr Phe Gln Leu Pro 165
 155 160 165
 Tyr Cys Val Pro Asn Glu Val Asp Tyr Tyr Phe Cys Asp Ile Pro 180
 170 175 180
 Val Met Leu Lys Leu Ala Cys Ala Asp Thr Ser Ala Leu Glu Met 195
 185 190 195
 Val Gly Phe Ile Ser Val Gly Leu Met Pro Leu Ser Cys Phe Leu 210
 200 205 210
 Leu Ile Leu Thr Ser Tyr Ser Gly Ile Val Phe Ser Ile Leu 220
 215 220

<210> 17
 <211> 326
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature

WO 01/98351

PCT/US01/19275

<223> Incyte ID No: 7477934CD1

```

<400>.17
Met Gly His Gln Asn His Thr Phe Ser Ser Asp Phe Ile Leu Leu
1 10 15
Gly Leu Phe Ser Ser Asn Lys Cys Gly Leu Leu Arg Gln Phe
20 25 30
Val Ile Phe Ile Met Ser Val Thr Glu Asn Thr Leu Met Ile Leu
35 40 45
Leu Ile Arg Ser Asp Ser Arg Leu His Thr Pro Met Tyr Phe Leu
50 55 60
Leu Ser His Leu Ser Leu Met Asp Ile Leu His Val Ser Asn Ile
65 70 75
Val Pro Lys Met Val Thr Asn Phe Leu Ser Gly Ser Arg Thr Ile
80 85 90
Ser Phe Ala Gly Cys Gly Phe Gln Val Phe Leu Ser Leu Thr Leu
95 100 105
Leu Gly Gly Glu Cys Leu Leu Leu Ala Ala Met Ser Cys Asp Arg
110 115 120
Tyr Val Ala Ile Cys His Pro Leu Arg Tyr Pro Ile Leu Met Lys
125 130 135
Glu Tyr Ala Ser Ala Leu Met Ala Gly Gly Ser Trp Leu Ile Gly
140 145 150
Val Phe Asn Ser Thr Val His Thr Ala Tyr Ala Leu Gln Phe Pro
155 160 165
Phe Cys Gly Ser Arg Ala Ile Asp His Phe Phe Cys Glu Val Pro
170 175 180
Ala Met Leu Lys Leu Ser Cys Ala Asp Thr Thr Arg Tyr Glu Arg
185 190 195
Gly Val Cys Val Ser Ala Val Ile Phe Leu Leu Ile Pro Phe Ser
200 205 210
Leu Ile Ser Ala Ser Tyr Gly Gln Ile Ile Leu Thr Val Leu Gln
215 220 225
Met Lys Ser Ser Glu Ala Arg Lys Lys Ser Phe Ser Thr Cys Ser
230 235 240
Phe His Met Ile Val Val Thr Met Tyr Tyr Gly Pro Phe Ile Phe
245 250 255
Thr Tyr Met Arg Pro Lys Ser Tyr His Thr Pro Gly Gln Asp Lys
260 265 270
Phe Leu Ala Ile Phe Tyr Thr Ile Leu Thr Pro Thr Leu Asn Pro
275 280 285
Phe Ile Tyr Ser Phe Arg Asn Lys Asp Val Leu Ala Val Met Lys
290 295 300
Asn Met Leu Lys Ser Asn Phe Leu His Lys Lys Met Asn Arg Lys
305 310 315
Ile Pro Glu Cys Val Phe Cys Leu Phe Leu Cys
320 325

```

```

<210> 18
<211> 2374
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 1714538CB1

```

```

<400> 18
gtcagtcagt ccaetggctc ccgcgcgcgc tctgtgtccg tcgctcggag ggtggaagcc 60
gggtctctgc gggccgctgc ccgatgact cctctctgcc tcaattgctc tgcctcctct 120
ggagacctgt acccaggggg tgcaaggaac cccatgctt gcaatggcag tgggcccagg 180
ggcactttg accctgagga ctggaacctg actgacgagg cactgagact caagtacctg 240
gggcccacgc agacagagct gttcatgccc atctgtgcca catacctgct gatctctgtg 300

```

WO 01/98351

PCT/US01/19275

```

gtggcgctg tgggcaatgg gctgacctg ctggatcatc tgcgccaaa ggccatgccc 360
acgctaccca actactacct cttcagcctg gcggtgctgg acctgctggt gctgctgggt 420
ggcctgcccc tggagctcta tggagatggt cacaaactacc ccttctctgct gggcgttggg 480
ggctgctatt tccgcaagct actgtttgag atgggtctgcc tggcctcagt gctcnaagctc 540
actgctctga gcgtggaacg ctatgtggcc gtgggtgcaac cactccagcc caggctcaatg 600
gtgaagcggg ccaatgtggc ccgagtgctt ggggcgctct ggggtcttgc catgctctgc 660
tcctgcccga aaccaagcct gcacggcabc gggcagctgc acgtgcccgt ccggggcccga 720
gtgccagact cagctgtttg catgtggtc ccgccacggg ccctctacaa catggtagt 780
cagcaacacc cgctgctctt cttctgctg ccaatggcca tcatgagcgt gctctgctg 840
ctgctggggc tgcactgctg gggggagggc ctgctgctca tgcaggaggc caaggcgagg 900
ggctctgcag cagcagggtc cagatacacc tgcaggctcc agcagcagca tccggggccg 960
ggacaagtga ccaagatgct gtttgcctg gtctgggtgt ttggcatctg ctgggcccc 1020
ttccacggcg accgctcat gtggagcgtc gtgtcacagt ggacagatgg cctgcaactg 1080
gcctccagc acgtgcaact catctccggc atcttcttct acctgggctc ggcggccaac 1140
ccgctgctct atagcctcat gtccagccgc tccgagaga cctccagga ggcctgtgc 1200
ctggggcct gctgcccctg cctcagacc ccgccacgct cccagcagct cagcaggatg 1260
accacaggga gcaccctgtg tgatgtgggc tccctgggca gctgggtcca ccccctggct 1320
gggacagatg gccacagagg cagcaagag acggatccat cctgggtgga gcttcaagt 1380
ggcttccctt ggaaggcca gaggtccacc tggagctggg gagaacatc tgcctctctc 1440
tgcaggatg ccttcagcta ctgtccctag ttcagcctag aaattctgac cagcactca 1500
gtttccctca gagggaaca ccaggaggag ggatccctga ctgctgagga ctcaactga 1560
ccagacgcca caccttggtc ttcttatctg tccactgcca ctccccagt tcaaatcct 1620
accctgaga aalatccacg ttagctgggg ctacagatc ctcccctggt ggactcctg 1680
ccaccactgc cagtttctga aacggctcca ctgggtccct actgtccctc ccagttcctg 1740
ttcaggttct ggcaggggccc cagggatcca ggggacctgg ttccaatctc agccctgctg 1800
tcaccacctt gtcattgcacc atcaagcata tcagctctacc tttctttttt tctgagacag 1860
agtctcactc tctcggccag gctggagtgc agtggcgcga ttttgaactg ctgcaacctc 1920
cgctccggg gttcaggcga ttctctctcc tcagcctccc gactagctgg gactaacagt 1980
gagccccagc atgcccagct aatttttttt agtttttagt agagacgggg ttccaccatg 2040
ttggccaggc tggctctgaa ctctgacct caggtgatcc ccgacctg gctcccaaa 2100
gtctcggat taagycatg agccaccaca ccggcccaat cagttccact ttctagctc 2160
tggttccttg cctgaaaaat gaagagggcg ctggcttctc acagtgtcat gctttggcac 2220
tttagctatg gttttctttc tgtgtgtgtg taagccactg ctataataa aaccaacaat 2280
accctcagac tgaaggggcg gaagtalta tctgcatett tatcaacccc aagccccact 2340
tcctccctga cctccccatg cctccccag cctc 2374

```

```

<210> 19
<211> 813
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<222> Incyte ID No: 3406743CB1

```

```

<400> 19
ctctctgggt cctcttcttc ctccgggaca aggatggagg atctcttttag cccctcaatt 60
ctcgcgcggg cgcaccaact tccgtgccc atcttctgctg gctggggtct caacctgacc 120
ttggggcaag gaggccctgc ctctgggccc ccacggccgc gcgtccgctt ggtgttctct 180
ggggtcatcc tgggtggggc ggtggcaggc aacaccacag tgcgtgtccg cctgtgccc 240
ggcggcgggc cctgggcggg cccaagcgt cgaagatgg acttctctgct ggtgcaactg 300
gccttgccgg acctgtaccg gtgcgggggc acggcgtctg cacagctgac ctgggaactg 360
ctggcgagcg ccgcgcggcg cacgggggac ctggcgtgccc gcttctctga gctgtgcaag 420
gcateccggc gggcgccctc ggcccacctc tgggtgctca tggccctcga gcgccggcgc 480
gcgccaggcg cgcacactctc cgcgccagcc tggccgggga tggcgtcgtg ccactggalc 540
ttcgctctcc tgcagcctg gcacgtgacg gtgtacgctt tctatgagcc cgtccggggc 600
ttcgtcggcg ctgtaagat catgggtgtc gcttggggcc acctactctc cgtctggtgg 660
cggcaccggc tgaaggcccc agcgggtaca gggcctggt cggcagccc aggtggagcc 720
cgtggcggca gcgcgatgcc ccggccaaag gtgcagagcc tgaagatgag ccagctgctg 780
ggctgctgtc tctggggctg cagctgccc tag 813

```

```

<210> 20
<211> 1542

```

WO 01/98351

PCT/US01/19275

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3485895CB1

<400> 20
tgaataaatt attcctgggc ttttgaatta ttaactcttg ctgcaggata tgcaaatagg 60
atatgcaaat aggcactctg gcacaacact ggttcagaag ctatattccag ctgctggcag 120
agttcctgtc aagggatcaa gtcttccaac agaattggtt tggtttaact cagcagaatt 180
tgtttgaaca ctacgacatg ctggggatca tggcatggaa tgcaactctg aaaaactggc 240
tggcagcaga gctgcccctg gaaaagtact acctttccat tttttatggg attgagttcg 300
ttgtgggagt ccttggaaat accattgttg ttacggccta catctctctc ctgaagaact 360
ggcaacgacg taatatttat ctctttaacc tctctgtctc tgacttagct tttctgtgca 420
ccctccocat gctgataagg agttatgcca atggaaactg gatataatgga gactgtgctc 480
gcataagcaa ccgatactgt ctctatgcca acctctaac cagcattctc ttctcactt 540
ttatcagcat agatcgtac tlgataatta agtatcctt ccagagaac ctcttgcaaa 600
agaanaagtt tgetatatta atctccttgg ccatttgggt ttatgtaacc tttaggttac 660
tcccatact tcccattata aatcctgtta taactgacaa tggcaccacc tgaatgatt 720
ttgcaagttc tggagaccoc aactacaacc tcatctacag catgtgtota acactgttg 780
ggttccttat tctcttttt gtgatgtgtt tcttttatta caagattgct ctcttccctaa 840
agcagaggaa taggcaggtt gctactgctc tggcccttga aaagcctctc aacttggta 900
tcatggcagt gtaaatctc tctgtgctt ttacaacctc tcaactcatg cggaaatgta 960
ggatcgctc acgctgggg agttggaagc agtatcagtg caactcagtc gtcataact 1020
aattttacat tgtgacacgg ccttggcct tctggaacag tgcactcaac cctgtctctc 1080
atcttctttt gggagatcac ttcagggaca tgcctgatga tcaactgaga cacaaactca 1140
aatcccttac atcctttagc agatgggctc atgaaactct actttcattc agagaaaagt 1200
gaggggcttg tgaaacagat tttctctacg atgaaactct aagccagttc cagtttgcct 1260
taactcatag acatcaatca gagagtgta cagcttaac ctgtatctaa agacaagttg 1320
taaccagagt atgtgaaaag aatgggacga caagaaatga ctggttctct cctctagaa 1380
ttgaaaggag ttgaactgcc ttatgtttgg gcatgtaact ccaaaatact aggtagtata 1440
aggctttctc aatcagtgca naaatggaag atataaaa caacaagttg tctgcatgg 1500
atcactggct agattgtaa aaaaaaaaa aagggcgccc gc 1542

<210> 21
<211> 1191
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476102CB1

<400> 21
atgggggatg agctggcacc ttgcccctgt ggcactacag cttggccggc cctgatccaag 60
ctcatcagca agacaccctg catgcccaca gcagccagca acacttctct gggcctgggg 120
gacctcaggg tgcaccagtc catgctgtac tggcttttcc ttcccctcaag cctgctggct 180
gcagccacac tggctgtcag ccccctgctg ctgggtgacca tctctgggaa ccaacggctg 240
cgacagagac cccactacct gctcccggct aacatcctgc tctcagacct ggctacatt 300
ctctccaca tgctcatctc ctccagcagc ctgggtggct gggagctggg ccgcatggcc 360
tgtggcaltc tcaactgatc tctctctgcc gctctcacca gcaactact gtctctcaac 420
gccattgtgc tgcacaccta cctggcagtc atccatccac tggcctacct ctctctcatg 480
tcccattggg ctgcccggaa ggcagtgccc ctcatctggc tgggtggcctg ctgcttcccc 540
scatctctta ttggctcag caagtggcag gatgccagc tggaggagca aggagctca 600
taactctac caacaagcat gggcaccag cgggatgtg gctctctggf catgtttacc 660
taacctcca ttctgtgggt tctgttctc tgcacagctc tcaattgcaaa ctggttctgg 720
agatctatg cagagcccaa gacttcaggc atctgggggc agggctatc cgggcccagg 780
ggcaccctgc tgatccactc agtctgac acattgtac tggacacag ggtgtgtctc 840
tcctggaca tgggtgtgac caggtaccac cacattgact ctgggactca cacatggctc 900
ctggcagcta acagtgaggt actcatgat ctctcccctg ccatctcac atacctgtac 960
ctgctccgct acggcagct gttgggcatg gtccggggcc acctccatc caggagggac 1020
caggccatct ttaccatttc ctgctgctgg gaaccactat tctacttttc tctgattttg 1080

WO 01/98351

PCT/US01/19275

accacctctag gaggatgat tattccattg tggtagaga ccacattttg tttatccatt 1140
catctgtcaa tggaccgttt gggttgcacc accotttggct attgtgaata a 1191

<210> 22
<211> 3360
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2432942CB1

<400> 22
cccacgcgto cggcagctga gacggcagcg gcagcttctc agggccggag ccagttcttg 60
gaggagactc tgcgcagggc atggatcaact gtgggtgccct tttcctgtgc ctgtgccttc 120
tgactttgca gaatgcaaca acagagacat gggagaact cctgagctac atggagaata 180
tgacaggtgtc cagggggcgg agctcagttt tttcctctcg tcaactccac cagctggagc 240
agatgtctact gaacaccagc ttcccaggct acaacctgac cttgcagaca cccaccatcc 300
agtlcttggc cttcaagctg agctgtgact tctctggcct ctgctgacc agtgccactc 360
tgaacgggtg gccccaagca ggaagtcagc atgcccgggg tcaagcaycc atgcaglttc 420
ccgcccagct gaccgggac gctgcaaga cccgcccag ggaagctggc ctactctgta 480
tctacttctc caaccccac ttttcaagg atgaaaaca ctcatctcg ctgaataact 540
acgtctggg gcccagctg agtcatggc acgtgaaca cctcaggat cctgtgaaca 600
tcagcttctg gcacaacca agcctggaag gctacacct gacctgtgc tctggaagg 660
agggagccag gaacagccc tggggggcct ggagccctga gggctgtct acagagcagc 720
cttcccactc tcaggtgctc tgcccgtgca accacctcac ctactttgct gttctcayc 780
aaactcccc agccctggct cctgcagagt tcttggcacc tcttaagctac atctccctgc 840
tgggctgcag catctccatc gtggcctcgc tgatcacagt cctgtctgac ttccatttca 900
ggaaagcagag tgaactcetta acacgcatec acatgaacct gcatgectcc gtgtgtctcc 960
tgaacatcgc ctctctctg agcccgcac tgcgaatgct tccgtgccc gggtcagcat 1020
gcacggctct gcccgtctgc ctgcactacg cgtctctcag ctgctcacc tggatggcca 1080
tcgagggctt caacctctac ctctctctcg ggcctgtcta caacctctac atccgcagat 1140
atgtttcaa gcttggtgtg ctgagctggg gggcccagc cctcctgtgt ctgcttccc 1200
tctctgtaa gactcggta taaggacct gcacaatccc cgtctctgac agctgggaga 1260
atggcacagg cticcagaac atgtccatg cctgggtcgg gagcccctg gtgcacagt 1320
tcctggtcat gggctacggc ggcctcacgt cctcttcaa cctggtgtgt ctggcctggg 1380
cgtctggac cctgcccagg ctgggggac ggggggatgc accaagtgc agggcctgccc 1440
atgacactgt cactgtctg ggcctcacgc tctgtctggg aaccactgg gccttggcct 1500
tcttttcttt tggcttctc ctgctgcccc agctgttctt cttaaccat ttaactcgc 1560
tgtacggttt ctctcttctc ctgtgttctt gctcccagcg gtgcccctca gaagcagagg 1620
ccaaggcaca gatagaggcc ttcagctctc ccaacaac acagtagtcc gggcctcctg 1680
gctcggaaatc ctacgctctc ctggcccaca gtgacctgag gctacggctc ctgctagaya 1740
gggtggcagg cctgtctctg gacccagag gccactgtga ccgccaaggg gcctttcca 1800
cttccagcg cctccaggc actgagggga aggcattgct ctacctctcc ctgacatttt 1860
gctccgggce agatccaac ttaactgggg cagcaaacct tgtctctgta cctgggcca 1920
gctcggcagg gatgtgggca gagcaccagg ctgggcatca gaaagccaag tttcaaggac 1980
tgtctttgag tctgtctgta tgacctggg cctgccactt ctccagagcc ctgagtaacc 2040
acagctgtga catgggggca agcagctttg tttcagccta acccagagc ttagtaaaaa 2100
ttgcataaga ccaggggaa gactgtcagc gtgggtggg aattcccgg gcctccacct 2160
gcttgctagg ggcaggatct cttcagggct gccctggaag cacctgctt gcccctgccc 2220
cttctccag gggagggcca gatggcatcc tggcttgggg cgggtgggac ctaccaggc 2280
tctgagactt tactggccta tgccagggc ctctttctct ttaactcctt aaattatgat 2340
gactccaaat ccaagcccac ccttcccaca gattggggag tcccgccgtt cccagagget 2400
cctctctcgg tgcctccaag acttccatag accatctgga ccagtagccc atcccagct 2460
ttcttgggg goagaggaaa acgttctctt ctctccagc tgaatcagct ggtcccagct 2520
gtcctggctg tttgggtgatt gggcaagatt gaatttgccc aggtaggcgt gagagtgtg 2580
gttttaact cgaagctcag gccatagttt cagagaatca ccttaacccc agacctctat 2640
gagacagtg tcatgaagc agtgcgttcc ccagaacgaa cactagggcg caccttggat 2700
ccacactcag aggcctctgg cgcacagact gactctagaa tctctcaaac actgtttg 2760
agccccatg caccagctgg agggccgcta actgagagc tggcctact gagtgaacca 2820
ttctctccag gaggaaaggc aagacacgct tacacggcca tttgtctctt tcccactgc 2880
ggggtgcac tttcctctt gggggctgca cccagacat agctggcacc agagcaggt 2940
gctcagctgg tgggtgctca gggcctgccc ccaggccact gggccgtttt gatgacctg 3000

WO 01/98351

PCT/US01/19275

```

aaggtcacag gcagaaaata ggagcaggat ttccctggg gaaaagtct cctgggacat 3060
ctctgctct tctgtacatt tctagatgca aataactcct tcaccaggca gtgagtggcg 3120
taggctctgg agccaggctg cctgggtccc aatgccagct ctgccacttg ctagtgtgga 3180
gactgtggac aaaccactca gccctctgtg gccctcagttt tctctattgt aaaaatagag 3240
ccatagtggt acctattttg aagactaagt aaaagaattc aaataaagag acctggcaca 3300
ggtaagtgc tcagtaaaaa aaaaaaaggg gggggccgga aaaaaggctc cgaaccggaa 3360

```

```

<210> 23
<211> 1660
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 4630911CB1

```

```

<400> 23
gtagatgaac aggaagcaac tggaggtaat acctatgatt tggglaatat ctctagataa 60
gggtaaaagg acttttcaca taggcagtca ctccacctct acgagtctca ttttttccaa 120
ctgcacaaat acagatttcc gatactttat ttatgcagtg acataactg tctctcttgt 180
gcccagcttc atagggaata tattagccct gtgggtattc tatggttata tgaagaagac 240
aaaacagact gtgatattta tgataaactt agccattgct gacttaetac aagttctttc 300
cttgccactg aggatctctt actacttgaa tcatgactgg ccatttgggc ctgtctcttg 360
catgttctgt ttctacctga agtatgtcaa catgtatgca agcatctact tcttggtctg 420
catcagtggt cgacgatttt ggtttctcat gtaccccttt cgcttccatg actgcaaaac 480
gaaatagcac ctgtacatca gcattgtctg ctggctgac atctgccttg cctgtgtact 540
ctttccactc ctacagaacca gtgatgatac cctgggcaat aggaccaaat gctttgtgga 600
tcttccctacc aggaatgtca acctggocca gtccgttgtt atgatgacca ttggcgagtt 660
gatgggtttt gtaactccgc ttctgattgt cctatatgtt acctggaaga cggttttatc 720
actgcaagat aaatatccca tggcccaaga ccttggagag aaacagaag ccttgaagat 780
gatctaaccc tgtgcagggg tattcctaatt ttgctttgca ccttataatt tcagttttcc 840
tttagatttc ctggtgaagt ccaatgaaat taagaactgc ctagccagaa ggggtattct 900
aatatttcab tctgtggcat tgtgtctgic tagtctgaaat tcatgtcttg acccagcat 960
atactacttt tccactaatg agttccgaag acggctttca agacaagatt tgcattgacg 1020
catccaactc catgcaaaat ccttltgtgag taaccataca gcttccacca tgcacctgca 1080
attatgctaa acaaaaaaac caactgaaat gtgacctgaa atgcaagtac atcagaacat 1140
atctgcaata cccaagccac aggaagaaac ttgcaaaaca acacagcttt tcagttctgc 1200
tctactctta tgcattgggg aattcaactc ttcaaacgag gacctatttg gagcattacg 1260
atccagcatt attgatgttg acatgctcat gtatgaattt ttcttcaagt ctgtaaatct 1320
taaaatatac aatttctgtg acatcctata aacatagcca ctcaactgta gtcagtactt 1380
ttactctgta acctcaggca caaaaagatt attagcttgg agtcaaccata aacatttagt 1440
ttgtttgcaa cagatactga gctctttatg tcagagggaaa tgtaagtgtc tttttatatt 1500
agtaatacaca gtttacatgc cattttcaba tgtttgggta tatttttagt ggatagatga 1560
tatattaccg tgtgtaaaga aaatgtaaac ataagatcat ttttatctct caagtgatga 1620
tactcttcag agtttgaaga ttaaaatagc tattttcttc 1660

```

```

<210> 24
<211> 2603
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472432CB1

```

```

<400> 24
ggaccagcag aaggaagaag tatgyagtta aagactgcag cgtgaactga gggatcccgg 60
accggcgcct tgcctcagag gatccagtc agatcccagg agagcccctc tgcacctctg 120
gaactcgtct cccatctaca aaagtggaag attggcccag attgccgtgc tctacaataa 180
gggtcatata ccactgcccc gctgcagcct gatctgagaa agcctctggc ccaccgcaat 240
ggccttctcg atgcacctgc tgggtctgctt cttcggaatg ggetcctggg tgaccatcaa 300
tggctctctg gttagactgc cctctgctgt gatggagctg cccgagggct ggtactctgc 360
ctcctactcc acggtggtca tccagctggc caaacctggg cccctctctg tcacctgtct 420

```

WO 01/98351

PCT/US01/19275

```

ccataccttc  oggcccagct  gectttccga  agtgcccatg  atcttcaccc  tgetgggctg  480
gggaaccgtc  acctgcatca  tctttgcctt  cctctgggat  atgacctctc  ggggtctgga  540
gggccaccac  agcabcgctc  tcttgctcct  caocttcttc  ctggccctgg  tggactgca  600
ctttcaagt  accttccctg  cgttcatgag  cggctgccc  acctactacc  tcaccacctt  660
cttttgggt  gaaggactca  ggggctcttt  gcccgccctg  gtggctcttg  cccagggtcc  720
cgttctcact  acctgctca  atgtcactya  gatatcagac  agcgtaccac  gccctgtacc  780
cagggaggag  actgacatgy  caagggagt  tccagagct  ttgggtctcg  cctccccgg  840
aatggaagca  cctctgtccc  acctggagag  cgcctacctt  cccgccact  tctaccctct  900
ggtcttctt  ctctccat  ccatcatgat  ggcctgtctg  ctctgtggct  cctttgtct  960
ccagcgtcaa  cccaggtgct  gggaggcttc  cgtggaagac  ctctcoaatg  accaggctac  1020
cctccactcc  atccggccgc  gggaaagaaa  tgacttgggc  cctgcagcga  cgggtggcac  1080
cagccagggc  caggggtatc  tagaggagaa  agcagccccc  tgctcccgg  cgcactctgg  1140
cttcatctat  acctgggtgg  cctctgtcaa  cgcgctcacc  aacggcatgc  tggcctctgt  1200
gcagacctac  tctgtcctgt  cctatggccc  agttgcttac  caoctggctg  ccacctctag  1260
cattgtggcc  aacctcttg  cctctgtggt  ctccatgttc  ctgcctaaca  ggtctctgct  1320
gttctgggg  gtctctctcg  tcttggggac  ctgctttggg  ggctacaaca  tggcctatgg  1380
gggtgatgag  cctgcccc  tcttgcaggg  ccactggggt  ggggaagtcc  tcattgtggc  1440
ctctggggtg  cttttcagcg  gctgctctag  ttactcaag  gtgatgctgg  gctgtgtct  1500
gcgcgacctc  agcgcagcg  cctctctgtg  gtgcccggcg  gcggtgcagc  tgggtctgct  1560
gtctggagg  gtgctatgt  tccctctggt  caacgtgctg  cggctctctc  cgtccggcga  1620
cttctgcaat  ctgactgttc  cagctlaggc  aggcgcgcga  cccgccccc  atgctcaag  1680
gacgcaactg  ggtccagag  agccaggtc  acagagcaag  gggcaggaac  agagagacag  1740
agctgagta  attgaaatca  gaacgaagt  gccactggg  gactgtgggg  aagatggcac  1800
ctggaaatgc  aaggtggggc  tctatcccca  actctgtgtc  acactacctg  tgacgacacg  1860
ctagatctc  cttgtcttg  actctcaaga  gaggactgat  ttgcagcctc  tagctggagg  1920
cagggccaag  ggtgttagaa  gggaaacagc  tgggacagcc  ggctgtccct  tcaggctgtg  1980
tgactctgg  aaagtcttt  ggtctctctg  tgcctgtttc  tcatgcatg  cagtggggat  2040
tccagtaagt  accaactacc  tcacagcat  ggcacggagg  caaaaggaaa  aagcagcccg  2100
catcaagcaa  gccctctgg  gccactgct  gatctgacag  tccatctgag  taacaagat  2160
ggcgtctgc  acaacctaga  agtgccaga  agggttgaga  cagcccccgt  cctctctccc  2220
tttgcccctc  agtctcacag  aggggtctct  acaagacaag  cagataacga  tagaatcttg  2280
ggcactctgg  ctttggatt  ctcaagtgtg  agggagctag  taacccacac  accctctcct  2340
gtcactcttc  ctggccaba  aagccacta  gttggagagt  agtaccctc  ctggagcag  2400
ggagagatga  ttgtctggtg  ggcctgggga  agccctctcc  ctgacctctc  gaaagtgaac  2460
tcccagacca  ggttggggac  cagacatgca  gagcccctgg  aagtattctc  tcaaatggag  2520
gcaacagagg  tgattgttat  ttgttttag  tttctgtttt  taattttttt  aaataaaggc  2580
atccctgct  ttaaaaaaaa  aaa

```

<210> 25
<211> 1119
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474977CB1

```

<400> 25
atggaggccg  ctagccttcc  agtggccacc  gccggcgttg  ccttgcct  gggaccggag  60
accagcagcg  ggaccccaag  ccgagagagg  ataactggtt  cgaccccgag  cggcgcctc  120
ctgcggggcc  gaggcccgcc  ctctctgttc  ttcacgtctc  tgggtgtgac  gctgctagt  180
ctgctgatcg  ccgccacttt  cctgtggaac  ctgctgtttc  cggtcaccat  cccgaggctc  240
cgtgctctcc  accgctgtcc  gcaataactg  gtggcctcga  cggcctctc  ggacgaaact  300
gtggcagcgc  tggcgtatgc  accgacctg  gcgagtgagc  tgtcgaccgg  gcgacgtcgg  360
ctgctggggc  ggagcctgtg  caccgtgtg  atctctctcc  acgtgctgtg  ctgcccggcc  420
ggcctcggga  acgtgggggc  catcgccctg  ggcgcgagc  gggccatcac  accgcaactg  480
cagcacacgc  tgcgaccccg  cagcccgccc  tcttctctca  tgatcggct  caccgggtg  540
cgttggcgc  tcatggcct  ggcgcgctg  ctcttggcc  gggcgagggt  gtgcgacct  600
cggctcagc  gctgcccagt  gggccggaaa  cctctctatg  cggctctctc  caccggggc  660
gctctccacc  tgcctctgg  cgtggtgct  ttgtctacc  ggaagatcta  cgaggcgcc  720
aagtttctgt  tggcccgcc  ccggagagct  gtgctgctg  tggccggcac  catgcaggt  780
aaggtaaagg  aagcactcga  tgaggctgaa  gtgtgttca  cggcacttg  caaagcaacg  840
gtgctctcc  aggtgagcgg  ggaactcctg  cgggagcaga  aggagaggcg  agcagccatg  900

```

WO 01/98351

PCT/US01/19275

```

atggtggaaa ttctgattgg cgtgtttgtg ctgtgctgga tccccttctt cctgacggaa 960
ctcatcagcc caactctgtc ctgcagcctg cccccatct ggaaaagcat atttctgtgg 1020
cttggctact ccaattcttt cttaaccccc ctgatttaca cagcttttaa caagaactac 1080
aacaatgctt tcaagagcct ctttactaag cagagatga 1119

```

```

<210> 26
<211> 1018
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7474848CB1

```

```

<400> 26
gggcaccagt ggaggttttc tgagcatgga tccaaccacc cgggcctggg gaacagaaa 60
tacaaccagt aatggaaatg accaagccct tcttctgctt tgtggcaagg agaccctgat 120
cccggtcttc ctgabccttt tcaatgccct ggtcgggctg gttagaaacg ggttctgtct 180
ctggctcttg ggtcttcgca tgcgcaggaa cgcctctctt gtctacgtcc tcaagcttgc 240
cggggccgac ttctctctcc tctgtctcca gattataaat tgcctgggtg acctcagtaa 300
cttctctgtg tccatctcca tcaattctcc tagcttcttc accactgtga tgacctgtgc 360
ctaccttgca ggcctgagca tgcctgagcag cgtcagcacc gagcctgcc tctcctctct 420
gtggcccatc tggatccgct gcccgccccc cagacacctg tcagcgtctg tctgtctctc 480
gctctggccc ctgtccctac tgcctgagcat cttggaaggg aagtctctgt gcttcttatt 540
tagtgatggt gactctggtt ggtgtcagac atttgatttc atcactgcag cgtggctgat 600
ttttttattc atggtttctc gtgggtccag tctggccctg ctggtcagga tctctctgtg 660
ctccagggtt ctgccactga ccaggctgta cctgaccatc ctgctcagag tctgtgtgtt 720
cctctctgct ggcctgcctt ttggcattca tgggttctca atattatgga tctggaagga 780
tctgtatgic ttattttgtc atattcatcc agtttcagtt gtctctcat ctcttaacag 840
cagtcaccaac cccatcattt acttctctgt gggctctttt aggaagcagt ggcggctgca 900
gcagccgcat ctcaagctgg ctctccagag ggcctctcag gacattgtgt aggtgataca 960
cagtgaagga tgcctcctgc agggcaccoc ggcagatgct agaagcagtc tgggttag 1018

```

```

<210> 27
<211> 2177
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7655614CB1

```

```

<400> 27
atggaggagg cgcagccgcc ccgcccacca gcgagcatgg ccttactggg cagccagcac 60
tcgagcggcc cctccgaggc cggcccacat ggcgggactt cctccggcgc caegcgggcc 120
gtgctctctc tcagcacctg gcgaccgcg gcgctgggga acctgagcga cgcaagcggc 180
ggcggcacag ctgcgctccc cggatggcgc gcccttgccc ggtccggggc agcgcggagc 240
gcggggcgcg cggtaaggcg gccctagagc ccgagggcgc cgcctctgct gtcgcaacga 300
gctgcagctg cggcccaggc gctctctctc ctgctcatct tctctctgct tagccttggc 360
aactgcgcgg tgatgggggt gattgtgaag caccggcagc tccgcacctg caccaacgcc 420
ttcatctctg cgtgtctctc atcggatctg ctccagcgcc tgcctctgct gccccggccc 480
ttcctggaac tcttcactcc gccggggggt tccggcctg ccccgccgcc ggggcccctg 540
cggcctctct gcgcgccag ccgcttcttc agctctgctc tccgcatcgt gtccaagctc 600
agcgtggcgc tcatctctgt ggaccgttac tggctatcgc tgcggccgcc gcgggagaa 660
atcggccgcc gccgcgcgct gcagctgctg gccggcgcct ggtgacggc cctgggcttc 720
tccttgccct gggagctgct cggggcggcc cgggaactcg cggcggcgca gacttccac 780
ggctgctctc accgacctc cccgacccc gcgagctgg gcgaggctt cagctggggg 840
ctggtggggg cctgctactc gctgaccttc ctgctcaggt gcttctgcca ctaccacatc 900
tgcagacggg tgcgctgtc gacgtgccc gtgcccggg tgaacacctc cgcgcgctg 960
ctgcccctct tcagcagagt gcgcagggcc accaccgtcc tcatcatgat cgtctctgct 1020
atctgctgct gggggcccta ctgcttctg gtgctgctgg ccgccgccg gcaggcccag 1080
accatgcagg ccccctgctc ctccagcgtg gtggcctgct ggtgacctg ggccaatggg 1140
gccatcaacc ctgtcatota cgcctatccc aatcccaaca tctctgctc cctagggcgc 1200

```

WO 01/98351

PCT/US01/19275

```

aaccgcgagg agggctaccg gactaggaat gtggacgott tccctgccag ccagggcccg 1260
ggctcgcag ccagaagccg cagtgcctt cgaaccgct atgccaccg gctggggccc 1320
tgcaacagga tgcctcttc caaccggcc agcggagtg caggggacgt ggccatgtg 1380
gcccgaaaa atccagttgt acttttctgc cgagagggac caccagagcc ggtgacgca 1440
gtgaccaaac agcctaatac cgaagctggg gataccagcc tctaagagcg ttggaatggc 1500
cagcttatga agcgaattt ccaactgcat tatttaalga tggaaagtc tggggggag 1560
ttgtggatc cataaagcca aacattlaaa gctagagccg ggggagcct accactttc 1620
ccaacacaa taagaagaaa tgcctcttc tcaaaagtc caaaggaat gtaaaatgca 1680
aaaatlaaa caatctfaa ccacataacc aagcattgtg aactgtaagt gccaaaatg 1740
acaaaataa cattactat aactgaaag ctcatatlat aggccaccac tbtgaacaa 1800
acaaaacatt gaatgcaacc agtattgttc aactacacag aa1ttcagga atgaatggag 1860
accttcagag ctgcttgaaa gccctctaa atcgctacca tccaacaaa ctttagctc 1920
tagagcttgc acctctgaca tgccttgggt gcttagattg caagtggaaa agtcttatcc 1980
ttacacacat tggtgaccct cctcacccca cccctaata acctttgtg aaaaatttta 2040
attcataacc catttaaaat cattttagga ttttgaaaag ctggttatta ttagataaaa 2100
ttcaaccatc tgaaggacaa atagaagtat gaacataact ttccagcaca ctgccccta 2160
ctagtgcca ggcctgt 2177

```

```

<210> 28
<211> 1632
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 6792419CB1

```

```

<400> 28
ttgatggaaa gggatgaatg aataaaagta cttgtctgat ggcagcagag accccgagca 60
aacggtggag gctacactgt ctggcattct cgcagcgttt cgtcagagcc ggaccgcct 120
gcagctcaag ggaggcgtgc tctctcccc gagcagcgtg gaaccagct gggttccgcc 180
tcccggagag gtggctccca tctctcgtc tgcctctggt ttgtcagatc cggagagctc 240
tgaactggcg gagggccc cggagcctt ctggagcagg tagcagcatg cagccgctc 300
caagctctgt cgaacggccc ctggctgccc tggctcttgc ctgagcagctg tccggatct 360
ggggagagga gagggcttc ccgctgaca gggccactc gctttgcaa accgacaga 420
taatgagcc acccactaag accttatggc caaaggcttc caagccagct ctggccggt 480
cgttggcacc tgcggagggt cctaaggag acaggacggc aggatctccg ccacgacca 540
tctcccctcc ccgctgccc gagcccatg agatcaagga gaacttcaaa tacatcaaca 600
cggttctgtc ctgccttctg ttctgtctgg ggatcaatcg gaactccaca cttctgagaa 660
ttatctcaaa gaacaagtgc atgcgaacg gtcccataat cttgtctgcc agctgtgctc 720
tggagacact gctgacatc gtcattgaca tccctatcaa tgtctacaag ctgctggcag 780
aggactggcc atttggagct gagatgtgta agctgtgccc tttcatacag aaagcctccg 840
tgggaatcac tgtgctgagt ctatgtgctc tgagtattga cagatatega gctgttctt 900
cttggagtag aattaaagya attggggctc caaaatggac agcagtagaa attgttttga 960
tttgggtggc ctctgagtt ctggctgccc ctggaagcat aggtttttag ataattaca 1020
tggactcaaa aggaagtat ctggaatct gcttcttca tcccgctcag aagaacgctt 1080
tcatgagatt ttacaagaca gcaaaagatt ggtggctgtt cagttctat tctgcttgc 1140
cattggccat cactgcattt tttatacac taatgacctg tgaatgtttg agaaagaaa 1200
gtggatgca gattgcttta aatgatcacc taagcagag acgggaagtg gcaaaaccg 1260
tcttttgcct ggtcctgtc tttgccctct gctggcttc cctcaccctc agcaggattc 1320
tgaagctcac tcttataat cagaatgat ccaatagatg tgaacttttg agctttctgt 1380
tggatttga ctatattgtt atcaacatgg cttcactgaa ttccctgatt aaccaatg 1440
ctctgtatlt ggtgagcaaa agattcaaaa actgctttaa gtcagctta tgcctgtgt 1500
gccagtcatt tgaagaaaaa cagtctctgg aggaaaaaga gtcgtgctta aagttcaaa 1560
ctaagtatca cgaatgatgac aacttccgtt ccagtaataa atacagctca tcttgaaga 1620
agaaaaaaa aa 1632

```

```

<210> 29
<211> 1458
<212> DNA
<213> Homo sapiens

```

<220>

WO 01/98351

PCT/US01/19275

<221> misc_feature
<223> Incyte ID No: 7474790CB1

```

<400> 29
atgccatca gctcggccca cggcabcac cgtccaacc tgcctggtat cttctcggc 60
gctctttcg tggcaacat agtgcggcg ctagtgttc aggcgaagc gcagctctg 120
caggtgacca acggttttat ctttaacct ctcgtcacc acctgctca gatttcgctc 180
gtggccccc ggggtgttc cactctgtg cctctctct gcccctcaa cggcacttc 240
tgaacggcc tggttagcct caccacctg ttcgccttc ccagcgtcaa caccattgc 300
gtgggtcag tggatcgcta cttgtccatc atccacctc tctctctcc gtccaagatg 360
accocagcc cgggttaact gctcctctat ggcacctgga ttgtggcat cctgcagagc 420
actcctccc tctacggctg gggccaggct gcctttgat agcgaatgc tctctctccc 480
atgctcggg gggccagccc cagctacact attctcagc tgggttcct catcgtcatt 540
ccactgatt tcatgattgc ctgctactcc gtgtgttct gtgcagccc gaggcagcat 600
gctctgctg acaatgtcaa gagacacagc ttggaagtgc gactcaagg ctgtgtggag 660
aatgaggatg aagagggagc agagaagaag gaggagtcc aggatgagag tgagtcttcg 720
cgccagcatg aaggtgaggt caaggccaag gaggccagaa tggaaagcaa ggacggcagc 780
ctgaaggcca aggaaggaag cccggggacc agtgcagta gtttagaggc cagggggcagc 840
gagagggtca gagagagcag cccggtggcc agcgacygca gcatggagg taaggaagc 900
agccccaag ttgaggagaa cagcatgaag gcagacaagg gtcgcacaga ggtcaaccag 960
tgcacattg acttgggtga agatgacatg gattttggtg aagacgacat caatttcagt 1020
gagatgagc tggaggcagt gaacatccc gagagctcc caccagtcg tcgtaacagc 1080
aacagcaacc ctctctgccc caggtgctac cagtcaaaag ctgctaaagt gatcttcact 1140
atcattttct cctatgtgct atcccctggg cctcactgct ttttagcagt cctggccgtg 1200
tgggtggatg tgaaaaccca ggtcaccagc tgggtgatca ccataatcat ctggtctttc 1260
ttcctgcagt gctgcatcca ccctatgtc tatggctaca tgcaacaagc cattaagaag 1320
gaatccagg acatgctgaa gaagtcttc tgcaaggaaa agccccgaa agaagatagc 1380
caccagacc tgcgggaac agaggtggg actgaaggca agattgtccc ttctcaagat 1440
tctgctactt ttccttga 1458

```

<210> 30
<211> 1015
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474816CB1

```

<400> 30
tgatatatat agtgtatata tacatacatg cacacataca ctatatatat caaggattta 60
tatgatagga ttaattaaga aaaaaattag tggaaataaa ataatgttta tgataatttt 120
ggccatagaa tatataatc agatgatgtg aagtaaaaa tgttttttat acttcabatt 180
ttgatgtaca aagtatgttt gcttttgtaa ttcagatgat taactttgac ttgtgtccc 240
atgaaaaatg cctttcattb ctaagctggt attggactct cagccaacac ttttctcctt 300
ctttctcggc tctctcctt tctgctttt tctggatcic aggccagagc gcacttaact 360
accagctctg catgtggccc tcatccacat ggtggtcctt ctcaccatgg tgtttctgct 420
tccacagctc tttgaatcac tgaattttca gaatgacttc aaatatgagg catctttcta 480
cctgagagag gtgacagagg acctctccat ttgaccacc tgcctcctgg gcatgtctga 540
ggctgtcaac atcagcccca gcatttctct gttgggtgag ttaaatgga aatccacaat 600
ttttactctc catttgttct catggtctct cagttttcct gttatagta gctgatctt 660
ttacactgtg gcttcttcca atgtgaccca gatcaattg catgtcagta aatactgttc 720
acttttccca ataaactcca taatcagagg actgttttc actctgctat tattcagaga 780
tgtttttctt aaacaaataa tgcgttctc aagtgtctac atgatgactc tcattcagga 840
actacagag atcctgtgac ctccacagcc ccagcctcta ctaaggatc ttgcaagag 900
caagagccat cagcacatcc tgcgtccggt gagtttctcg gtyggcatgt acaagatgga 960
cttcatcctc tcaacctcct caacgttgcc atgggcatac gaccgtggtg tctag 1015

```

<210> 31
<211> 2781
<212> DNA
<213> Homo sapiens

WO 01/98351

PCT/US01/19275

<220>
<221> misc_feature
<223> Incyte ID No: 7476172CB1

```

<400> 31
atggacttct taattatact aattacotgc tttgtgatta ttcttgetac tteacagcct 60
tgcagagacc ctgatgactt tgtgctgccc acttctccgg gacatataat aattggaggt 120
ttgtttgcta ttoatgaaaa aatgttgccc tcaagaagact ctcccagagc accacaatc 180
caggagtggt ttggctttga aatalcagtt ttcttcaaa ctottggcat gatcacnagc 240
attgagatga tcaacaattc aacactotta cctggagatc aactggggta tgaatctat 300
gacacttgta cagaagtccac agtggcaatg gcagccaactc tgaggtttct ttccaatlc 360
aactgtccca gagaaactgt ggagtttaag tgtgactatt ccagctacat gccaaagatt 420
aaggctgtca taggttctgg gtactcagaa ataactatgg ctgtctccag gatgttgaat 480
ttacagctca tgcacacaggt gggttatgaa tcaactgcag aaactctgag tgacaaaatt 540
cgctttcctt cttttttacg gactgtgccc agtgacttcc atcaaatiaa agcaatggct 600
caactgattc agaaatctgg ttggaactgg attggcatca taaccacaga tgatgactat 660
ggacgattgg clcttaacac ttttataatt caggctgaag caaataaact gtgcatagcc 720
ttcaaaaggg ttcttccagc ctttcttcca gataatccca tgaagtcag aatcaatcgg 780
accttgaaqa aatcatcttt agaagcccag gttaatgcca ttgtggtatt tctgaggcaa 840
ttccatgttt ttgatctctt caataaagcc attgaaatga atataaataa gatgtgatt 900
gctagtgata atbggtcaac tgcacacaag attaccacca ttccatattg taaaagatt 960
ggcaaaagtg tagggtttgc ctttagaaga ggaatataat cctcttcca ttcttctct 1020
caaaatctgc acttgcctcc cagtgcacgt cacaactctc tacatgaata tgcactgcat 1080
ttatctgctc gcgcataatg caaggacact gatttgagtc aatgcataat caatcattct 1140
caaaagactt tggcctacaa ggtaacaagc gctatagaaa ggaacttctc catgagaaat 1200
gacttctctc gggactatgc tgagccagga ctcaatcata gtaltcagct tgcagtttt 1260
ggccttggtt atgcccattg ggtctgtgt caagctcgtg actgcagaaa ccccaacgcc 1320
tttcaacctc gggagttact tgggtgcta aaaaatgtga cattcaactg tggatggaat 1380
tcatttcaat ttgatgtcca cggggattta aatactggat atgatgttgt gctctggaag 1440
ggactcaatg gacacatgac tgtcactaag atggcagaat atgacctaca gaatgatgc 1500
ttcacatccc cagatcagga acaaaaaaat gagtccagga atcttaagca aactcaatct 1560
aaatgtcca agaatgcag tcttgggcaa atgaaagaaa ctacaagagc tcaaacatc 1620
tgbttctatg aatgtcagaa ctgtctgtaa aatcattaca cbaatcagac agatagcct 1680
cactgctt tttatgcaaaa caaaactcac tgggcccctg ttaggagcac tatgtgctt 1740
gaaaagaagc tggaaatctc caactggaat gactccttg ccatcctact cctgattctc 1800
tcctactggc gaatcatatt tgttctggtt gttggcataa tatttaacag aaacctgaac 1860
acacctgttg tgaatcctc cgggggatta agagctgct atgtgatct tctctgcat 1920
ttctcaatt ttgcccagac gagctttttc atttggagaa cacaaagact cacatgaaa 1980
accaggcaga caatgtttgg agtgagcttt actctttgca tctcctgcat tttgacgaag 2040
tctctgaaaa ttttggtagc ctctcagctt gatcccaaat tacagaaatt tctgaagtc 2100
ctctatagac gpatccttat tatcttcaact tgcacgggca tccaggttgt catttgaca 2160
ctctggctaa tctttgcagc acctactgta gaggtgaaat tctccttgc cagagtcac 2220
atcctggagt tggaggagg atccatactt gcatttggca ccatgctggg ctacatggc 2280
atcctggcct tcaattgctt cataattgct tcaaaagca aatatgagaa ttcaaatgaa 2340
gccaaitca ttacatttgg catgctcatt tactcatag ctggatac acatctcct 2400
atctatgta ccaatattgg caaataatgta ccagctgtgg agattattgt catattata 2460
tctaaactatg gaatcctgta ttgcacatc atccccaaal gctatgttat tatttgaag 2520
caagagatta acscaaagtc tgcctttctc aagatgatct acagttatct tccccatg 2580
gtgagcagca ttgcccctgag tctgtcttca ctggactcca tgagcggcaa tgtcaaatg 2640
accaatccca gctctagtgg caagtctgca acctggcaga aaagcaaaqa tcttcaggca 2700
caagcatttg cacacatag cagggaaaaat gccacaagtg tatctaaaaa tttgcctga 2760
aaaagaatgt caagtatatg a

```

<210> 32
<211> 1267
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472141CB1

<400> 32

WO 01/98351

PCT/US01/19275

```

ccccccct tcccccgga ggggggaaa accatgtaaa ctctcacagc tctaaaaaat 60
ccctttatgt cccttggctc gcatctccta tgaagaagc ttaataatag aaacgatccc 120
caattttacaa tgcttgaalc gcaactacaca ccacaaattc acgcttaata tttttatggg 180
ctgataggca aatatctaaa acaltcacaa gaacttogcc cgcctttaga aactactctc 240
atgaactatc ttccagctgg catcaacctta ttccctactc aagacttatg gacaagactg 300
ttagtattcc acaactatc ctatctcttc tttttaaag ttaattgaac acacataaaf 360
cactgaattt cctctcttgg gggtagacaga catcaagaa ttaaaacctc tctcttttgt 420
catgttcctt gccatgtact ttataaatgt gtttgggaat ggagccatca tgatgattgt 480
catcttagat ccaagaactct actcacctat gtatttcttc ctgggaacc tagcatgctt 540
agatatactgc ttctccactg taacagtgc aaaaatgctg gagaactctt tctctacaag 600
caagaqaatt tccttcttgg gatgcataac tcaacttcat ttcttccact tcttgggaag 660
cacagaagcc ctgtctgtga cagtgatggc atttgaccgc tttgtggcta tctgcagacc 720
actccactac cctgcatca tgaatcgcca gctctgtatt cacatgactg ttactatctg 780
gaccattggc tttttccatg cctctcttca ctctgtaatg acatctcgtt tgagctctctg 840
tggttccaat catatccatc atttctcttg tgatgtaag ccattactgg atctggcctg 900
tggaaacact gagctcaacc ttggctgct caatacagtc acaggacca tgcctctcac 960
ttccttttat ctgatalttc tctcttattt ctacatctc accaatcttc tctcaagac 1020
ccgtctctgc agcatgctcc acaaaactct gctacactct gccctccact tcatggctgt 1080
tgttctgttc catgctcctg ttctctctac ctacatcctg ctgctctcag gcaactctct 1140
ggatccaggat acaatcactg ccactcatga cagtgtggtc accctgctc tcaatccact 1200
cstgtacacc ttgagaaca aggaagttag gagtgcattg aatgaaagg tgagaagctc 1260
cctttga

```

<210> 33
 <211> 1559
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472137CB1

```

<400> 33
atagtttggc tgacatttta gcagatactt cataaacttt actataatgc taggatattg 60
tgttaccgat gtgttataat aagtatgatt tccattggca gatgatgaa ctaagtctat 120
gaaagagctc atgtaaatgt ttaacaagt atagctaata aatggagagg ccaggatca 180
lahtaagttt tctgttttct aacctaaata tttttctgtt acatlgcaga gggcaatgaa 240
cagtggagtt ttcttatgta aagacatcac aagagttcaa atagtaata tgcctaaaaa 300
gaataaatcc atggtacata ggccttaaac ttaagctaaa gtgtttttaa tatgcctcta 360
aacttccatc cctgtagaag acaagaatct gcctttacaa agcatccctt ccatgagctg 420
aatagactga ctctaaaata ggcatttctt tattatctga ttttttttta catggatctc 480
tctcttgcct ttcataagaa ggcaggttgt tcttgttaa aaacatagcc aaagcactca 540
tcagcaaaaa gcttgagaac tbgacattgg tbttagact tgggtcctg ggtccctttt 600
caaggctgga tgttttcaaa gccatgagga atttctcgtt ggtgtccgaa ttcactctgc 660
tgggcatccc tcaacaggag ggtctggaga ctattctgtt ggtctgttt ttgtcctctt 720
acatcttccac ccttatgggg aacctgctca tcttctggcc tattgtctcc tctgctcggc 780
tccacagcc catgtacttc ttctgtgca agctgtctgt ttttgacctt ttttccctt 840
ctgtgagttc ccttaagatg ctgtgctatc tttcagggaa cagccagacc atctctctatg 900
caggctgtgc atcccagctc ttctcttacc atttccggg ctgactgag tgtttctctg 960
acaggtgatg ggcctacgac cgtttgttg ccattgtgca cctctacgc tacaccataa 1020
tcatgagcca cagagcatgt atcatcctag ccattggggc ctcatctttt ggtgcatctc 1080
agccacactt tctgaccact ctcaacttcc aattgcctta ctgtgtcccc aatgaggtgg 1140
actattatct ctgtgatatc ccagtcacgc tgaagctggc ttgtgcagat acctagagcc 1200
tggagatggt ggggttccatc agtgtgggco tcatgcccct cagctgttcc ctctctatcc 1260
tcacctccta cagtggcatc gtcttctcca tctttagatc ctgctctgcc gaggccagc 1320
gccgtgcctt ctccacctgc agcgcaccac tcaccgccat cctgcttttt taactgcccag 1380
tggctctcat ttaactgagg cctaccacaa gccctgggtt ggtatgcaact gttcaanttc 1440
tgaataacct ggtcaccccc atgctgaacc ccttaactca cagtctcagg aataaggagg 1500
tgaattatc actaaggaag gtcttatato agctgggctt ccttctctag cagttagt 1559

```

<210> 34
 <211> 981
 <212> DNA

WO 01/98351

PCT/US01/19275

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7477934CB1

<400> 34

```

atgggccatc agaatcacac ttacagcagt gatttcatac ttttgggatt gttctcttct 60
aaccaagtggt gctctctctt tagacaattt gtcattttca ttatgagtg aacagaaaat 120
acgtccatga tctctctcat tgcagtgac tcccgactcc acactccaaat gattttcttg 180
ctcagccatc tctctctaat ggatatcttg catgtttcca acatcgttcc caaaatggtc 240
actaaccttc tgcagggcag cagaactatt tcaatttcag gttgtgggtt ccaggtattt 300
ctgtccctca cctctctggg tggtagtgac ctctctctgg ctgcaatgtc ctgtgatcgc 360
tatgttgcta tctgtcaccg gctgcgctat ccgattttta tgaaggagta tgccagcgct 420
ctcatggctg gaggtctctg gctcattggg gttttcaact ccacagttca cacagcttat 480
gcaactgcat ttccctcttg tggctctagg gcaattgac acttctctg tgaagtcctc 540
ccatgttga agtgtctctg tgcagacaca acacgctatg aacgaggggt ttgtgtaagt 600
gctgtgatct tctgtctgat cctctctctc tgsatctctg ctctctatgg ccaaatattt 660
cttactgtcc tccagatgaa atcattcagag gcaaggaaaa agtctatttc caactgtctc 720
ttccacatga ttgtgtctac gatgactat gggccattta tttttacata tatgagacct 780
aaatcattcc acactccagg ccagataaag ttctctgcaa tattctctac gatctcaca 840
cccaactaca accctttcat ctacagcttt aggaataaag atgtttctgg ggtgatgaaa 900
aaatgtctca aaagtaactt tctgcacaaa aaaatgaata ggaaattcc tgaatgtgtg 960
ttctgtctat ttctatgta a

```

<210> 35

<211> 2177

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7655614CB1.comp

<400> 35

```

acgagccttg acactagtac ggcgcagtg gctggaagt tatgttcata cttctatttg 60
tcttcagat ggttgaattt tatctaataa taaccagctt tcaaaaatcc taaaatgatt 120
ttaaattgggt tatgaattaa aaattttcca caaagtttat taggggttgg ggtgagggag 180
tgcccctaa gttgtaagga taagactttt ccacttgcaa tctaagcacc caaagcatgt 240
cagagggtga agctctagaa gctaaagttt gtttggatgg tagcattta gaggggcttt 300
caagcagctc tgaaggtctc cactcattcc tgaattctg tgtagtga caactctgtt 360
tgcattcaat gttttgtttg ttccacagtg tggctcctata atatgagett ttcagtata 420
gtgaatgta tttttgcat ttttggcaet taccagttcac aabgcttggg tatgtggtt 480
aagattgttt taatttttgc attttacatt ccttttggca cttttgaaga agggacattg 540
tcttttatgt tgtttgggga aagtgttaag cctccccctg ctctagcttt aaatgttttg 600
ctttatgaaa tcccaactc tccccagaa tcttccatca ttaaataatg cgagtggaaa 660
ttgctctca taagctggcc attccaaccg tcttagagggc tggatcccc agctctggat 720
ttagctgttt tggctactgc cgtcaccggc tctgtggttc cctctcgga gaaaagtaca 780
actggatttt tgcggcccca catggccaag tcccctgcca ctccgctggc cgggttggaa 840
gaggacatcc tgttgaggc cccagccgg ttggcatagc ggtttcgaag gcgactcgg 900
cttctgctct gcagaccggc gccctggctg ggcaggaagc cgtccacatt cctagtccgg 960
tagccctctc cgcggttgcg ccttaggagc atcgaatgt tgggattgcg gatggcgtag 1020
atgacaggtt tgaatgcccc attggcccag gtcagaccaga cggccaccac gctgaggagc 1080
gagggggcct gcaatggtctg ggcctgccc ggcggcgcca gcagaccag gaagcagtag 1140
ggcccccagc agcagatgac gaagacgac atgatgagga cgggtgtggc cgtgcgacc 1200
tcgttgaaga agcgcagcac ggcgcgtag tgttccaccg gccgcacgca cactgtccag 1260
agcgcaccg tcttgcagat gttgtgtg cagaagcaca tgaagcagaa ggcagcagg 1320
tagcaggcca ccaccagccc cactgtgaag gccgcccaca gctgcccggg gtcaggagg 1380
gtcccgtaga gccagcctg gaagctctgc gccgcgcga gttcccggg cgcctcagc 1440
agctcccaag gcaagagaaa gccacggccc gtcagccagg cgcctccagc cagctcagc 1500
gcgcggcgcc gccagatctt ctcccggcc ggcgcagca tagcgcagta acggtccaac 1560
gagatgagcg ccacgtctgag cgtggcacgc atgcccgaag acgagctgaa gaagcggctg 1620
gcggcgca gaaccgcgcca gggccccgc ggcggcgag gcgcccgaac cccggcgga 1680

```

WO 01/98351

PCT/US01/19275

```
gtgaagaggt ccaggaaggc ggcgggcagg cagagcagcg ccgtgagcag atccgatagg 1740
gacagcgaca ggatgaaggc gttggtgacg gtgcggagct gccggtgctt cacaatcacc 1800
cccatcaccg cgcagttgcc aaggctagac agcaggaaga ttagcaggag gacgagcgcc 1860
tgggcccgca ctgcagctcc gtgcgacagc agcggcgccg cctccgggcc tagcgccgc 1920
ctaacggcg ccccgcttc ccgctctcc ccgactctgc caagcccgcc gccaccgga 1980
gggcagctg tgcgctcc gcttgcttg ctcaagttcc ccagcgccg ggtcgccag 2040
gtgctgaagg agagcagcg cgcctggcc gcgaggaag tcccgccag tggcgccgc 2100
gcccagggg ccccgagtg ctggctgcc agtaaggcca tgctcctgg tggcggggc 2160
ggctgcgct cctccat 2177
```

【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
27 December 2001 (27.12.2001)

PCT

(10) International Publication Number
WO 01/098351 A3(51) International Patent Classification: C12N 15/12,
C07K 14/705, C12N 5/10, C07K 16/28, C12Q 1/68,
G01N 33/50, A61K 9/00, 39/395

(21) International Application Number: PCT/US01/19275

(22) International Filing Date: 15 June 2001 (15.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/212,483 16 June 2000 (16.06.2000) US
60/213,954 22 June 2000 (22.06.2000) US
60/215,209 29 June 2000 (29.06.2000) US
60/216,595 7 July 2000 (07.07.2000) US
60/218,936 14 July 2000 (14.07.2000) US
60/219,154 19 July 2000 (19.07.2000) US
60/220,141 21 July 2000 (21.07.2000) US(71) Applicant (for all designated States except US): INCYTE
GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo
Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): LAL, Preeti
[IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US).
BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road,
San Leandro, CA 94577 (US). HAFALLA, April, J., A.
[US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054
(US). NGUYEN, Danniell, B. [US/US]; 1403 Ridgewood
Drive, San Jose, CA 95118 (US). GANDHI, Ameena,
R. [US/US]; 837 Roble Avenue, #1, Menlo Park, CA
94025 (US). KALLICK, Deborah, A. [US/US]; 900
Olive Street, Menlo Park, CA 94025 (US). GRIFFIN,
Jennifer, A. [US/US]; 33691 Mello Way, Fremont, CA
94555 (US). YUE, Henry [US/US]; 826 Lois Avenue,
Sunnyvale, CA 94087 (US). KHAN, Farrak, A. [IN/US];
333 Escuela Avenue, #221, Mountain View, CA 94040
(US). PATTERSON, Chandra [US/US]; 490 Sherwood
Way #1, Menlo Park, CA 94025 (US). LU, Dyung,
Aina, M. [US/US]; 233 Cay Drive, San Jose, CA 95123
(US). TRIBOULEY, Catherine, M. [FR/US]; 1121
Tennessee Street, #5, San Francisco, CA 94107 (US).
LU, Yau [CN/US]; 3885 Corina Way, Palo Alto, CA94303 (US). WALIA, Narinder, K. [US/US]; 890 Davis
Street, #205, San Leandro, CA 94577 (US). GRAUL,
Richard [US/US]; 682-29th Avenue, San Francisco, CA
94121 (US). YAO, Monique, G. [US/US]; 111 Frederick
Court, Mountain View, CA 94043 (US). YANG, Junming
[CN/US]; 7125 Baric Lane, San Jose, CA 95129 (US).
RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird
Circle, Fremont, CA 94555 (US). AU-YOUNG, Janice
[US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005
(US). ELLIOTT, Vicki, S. [US/US]; 3770 Polton Place
Way, San Jose, CA 95121 (US). HERNANDEZ, Roberto
[GB/GB]; 160 Pine Tree Avenue, Canterbury Kent, CT2
7SZ (GB). WALSH, Roderick, T. [IE/GB]; 8 Boundary
Court, St. Lawrence Road, Canterbury, Kent CT1 3EZ
(GB). BOROWSKY, Mark, L. [US/US]; 122 Orchard
Avenue, Redwood City, CA 94061 (US). THORNTON,
Michael [US/US]; 9 Medway Road, Woodside, CA
94062-2612 (US). HE, Ann [CN/US]; 4601 Catalina
Drive, San Jose, CA 95129 (US).(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics,
Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DO, DZ, ES, FI, GB, GD, GE, GR, 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,
TH, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.(84) Designated States (regional): ARIPO patent (GI, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BI, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

(88) Date of publication of the international search report:
9 January 2003For 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.

(54) Title: G-PROTEIN COUPLED RECEPTORS

(57) Abstract: The invention provides human G-protein coupled receptors (GCREC) and polynucleotides which identify and encode GCREC. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of GCREC.



WO 01/098351 A3

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT		International Application No. PCT/US 01/19275
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N5/12 C07K14/705 C12N5/10 C07K16/28 C12Q1/68 G01N33/50 A61K9/00 A61K39/395		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 7 C12N C07K C12Q G01N A61K		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data bases consulted during the international search (name of data base and, where pertinent, search terms used) EPO-Internal, SEQUENCE SEARCH, PAJ, WPI Data, BIOSIS		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 00 02919 A (MCKEE KAREN KULJU ; LIU QINGYUN (US); HOWARD ANDREW D (US); MERCK &) 20 January 2000 (2000-01-20) see SEQID 10 + 3 the whole document ---	1-7, 9, 12-14, 16, 17, 19, 20, 22, 23, 26, 45, 62
X	WO 00 22129 A (ARENA PHARMACEUTICALS INC ; LIAW CHEN W (US); BEHAN DOMINIC P (US);) 20 April 2000 (2000-04-20) see SEQID114 + 113 the whole document --- -/--	1-7, 9, 12-14, 16, 17, 19, 20, 22, 23, 26, 45, 62
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents :		
"A" document defining the prior art which is not considered to be of particular relevance		
"E" earlier document but published on or after the international filing date		
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)		
"O" document relating to an oral disclosure, use, exhibition or other means		
"P" document published prior to the international filing date but later than the priority date claimed		
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention		
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone		
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.		
"Z" document member of the same patent family		
Date of the actual completion of the international search	Date of mailing of the international search report	
23 April 2002	26. 07. 2002	
Name and mailing address of the ISA European Patent Office, P.O. Box 5818 Patentstrasse 2 NL - 6280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 851 epo nl, Fax: (+31-70) 340-3010	Authorized officer Holtorf, S	

Form PCT/IB/200210 (patent sheet) (July 1999)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/19275

C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL SEQUENCE LIBRARY [Online] 14 December 1999 (1999-12-14) SUN, H., ET AL.: "the sequence of the Homo sapiens BAC clone RP11-56215" XP002197179 accession no. AC017104 ---	1-5,11, 12,45,62
A	LIU M ET AL: "GPR56, a novel secretin-like human G-protein-coupled receptor gene" GENOMICS, ACADEMIC PRESS, SAN DIEGO, US, vol. 55, no. 3, 1 February 1999 (1999-02-01), pages 296-305, XP002137722 ISSN: 0888-7543 the whole document -& DATABASE EMBL SEQUENCE LIBRARY [Online] 17 May 1999 (1999-05-17) LIU, M., ET AL.: "GPR56, a novel secretin-like human G-protein-coupled receptor gene - Homo sapiens G-protein-coupled receptor (GPR56) mRNA, complete cds" XP002194336 accession no. AF106858 ---	
A	WO 00 15793 A (INCYTE PHARMA INC ;CORLEY NEIL C (US); BANDMAN OLGA (US); BAUGHN N) 23 March 2000 (2000-03-23) see Table 2/SEQ1D6 the whole document ---	
A	WO 99 63087 A (HODG MARTIN R ;GLUCKSMANN MARIA ALEXANDRA (US); MILLENNIUM PHARM 1) 9 December 1999 (1999-12-09) figure 2 ---	
A	OLIVEIRA L ET AL: "A COMMON MOTIF IN G-PROTEIN-COUPLED SEVEN TRANSMEMBRANE HELIX RECEPTORS" JOURNAL OF COMPUTER-AIDED MOLECULAR DESIGN, ESCOM SCIENCE PUBLISHERS BV, XX, vol. 7, no. 6, 1 December 1993 (1993-12-01), pages 649-658, XP002050853 ISSN: 0920-654X the whole document ---	
	-/--	

Form PCT/ISA216 (continuation of record sheet) (July 1992)

INTERNATIONAL SEARCH REPORT		International Application No PCT/US 01/19275
C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	NG GORDON Y K ET AL: "Cloning of a novel G-protein-coupled receptor GPR 51 resembling GABA _B receptors expressed predominantly in nervous tissues and mapped proximal to the hereditary sensory neuropathy type 1 locus on chromosome 9." GENOMICS, vol. 56, no. 3, 15 March 1999 (1999-03-15), pages 288-295, XP001075001 ISSN: 0888-7543	
A	-& DATABASE ENBL SEQUENCE LIBRARY [Online] 4 January 1999 (1999-01-04) NG, G.Y.K., ET AL: " Homo sapiens orphan G protein-coupled receptor HG20 (HG20) mRNA, complete cds." XP002197180 accession no. AF069755 ---	
P,X	DATABASE TREMBL DATABASE [Online] 1 March 2001 (2001-03-01) RADDATZ, R., ET AL.: "Identification and Characterization of Two Neurotrophin U Receptors Differentially Expressed in Peripheral Tissues and the Central Nervous System" XP002197181 accession no. Q9HB89	1,12,45, 62
P,X	-& DATABASE ENBL SEQUENCE LIBRARY [Online] 25 October 2000 (2000-10-25) RADDATZ, R., ET AL.: "Identification and Characterization of Two Neurotrophin U Receptors Differentially Expressed in Peripheral Tissues and the Central Nervous System - Homo sapiens neurotrophin U receptor 1 (NUR1) mRNA, complete cds." XP002197182 accession no. AF272362 ---	1,12,45, 62
P,X	RADDATZ RITA ET AL: "Identification and characterization of two neurotrophin U receptors differentially expressed in peripheral tissues and the central nervous system" JOURNAL OF BIOLOGICAL CHEMISTRY, AMERICAN SOCIETY OF BIOLOGICAL CHEMISTS, BALTIMORE, MD, US, vol. 275, no. 42, 29 October 2000 (2000-10-20), pages 32452-32459, XP002163227 ISSN: 0021-9258 the whole document --- -/--	1-5,11, 12,15, 45,62

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

INTERNATIONAL SEARCH REPORT		Int. Application No. PCT/US 01/19275
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	WO 01 40797 A (HOSOYA MASAKI ;KAWAMATA YUJI (JP); FUKUSUMI SHOJI (JP); HINUMA SHU) 7 June 2001 (2001-06-07) see abstract ---	1-7,9, 12-14, 16,17, 19,20, 22,23, 26,45,62
E	WO 01 44297 A (SYNAPTIC PHARMA CORP) 21 June 2001 (2001-06-21) page 116 -page 134 ---	1-5,11, 12,15, 19,22, 25,45,62
E	WO 01 75093 A (CHEN RUI HONG ;HYSEQ INC (US); WANG DUNRUI (US); WANG JIAN RUE (US)) 11 October 2001 (2001-10-11) the whole document -----	1-5,11, 12,45,62

Form PCT/ISA210 (continuation of second sheet) (July 1999)

International Application No. PCT/US 01/19275

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 219

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

1. Claims: 1-44 partially, 45, 62 completely

1. Claims: 45,62 completely; 1-44 partially

An isolated polypeptide sequence of SEQ ID Nos. 1; an isolated polynucleotide sequence of SEQ ID No. 18 encoding said polypeptide; a cell transformed with said polynucleotide sequence; a transgenic organism comprising said polynucleotide sequence; a method for producing said polypeptide encoded by said polynucleotide sequence; an isolated polyclonal and/or monoclonal antibody which binds to said polypeptide sequence; a method for detecting a target polynucleotide sequence; a composition comprising said polypeptide and a pharmaceutical acceptable excipient; a method for screening a compound for effectiveness as an agonist and/or antagonist of said polypeptide; a method for assessing toxicity of a test compound using said polynucleotide sequence; a diagnostic test for a condition or disease using said antibody; a method for detecting said polypeptide in a sample.

2. Claims: 1-44

2. Claims: 46,63 completely; 1-44 partially
Idem as invention 1 but limited to SEQ ID Nos. 2 and 19.

3. Claims: 47,64 completely; 1-44 partially
Idem as invention one but limited to SEQ ID Nos. 3 and 20.

4. Claims: 48,65 completely; 1-44 partially
Idem as invention 1 but limited to SEQ ID Nos. 4 and 21.

5. Claims: 49,66 completely; 1-44 partially
Idem as invention 1 but limited to SEQ ID Nos. 5 and 22.

6. Claims: 50,67 completely; 1-44 partially
Idem as invention 1 but limited to SEQ ID Nos. 6 and 23.

7. Claims: 51,68 completely; 1-44 partially
Idem as invention 1 but limited to SEQ ID Nos. 7 and 24.

8. Claims: 52,69 completely; 1-44 partially
Idem as invention 1 but limited to SEQ ID Nos. 8 and 25.

9. Claims: 52,70 completely; 1-44 partially
Idem as invention 1 but limited to SEQ ID Nos. 9 and 26.

10. Claims: 52,71 completely; 1-44 partially
Idem as invention 1 but limited to SEQ ID Nos. 10 and 27.

International Application No. PCT/US 01/19275

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

11. Claims: 52,72 completely; 1-44 partially
Idem as invention 1 but limited to SEQ ID Nos. 11 and 28.
12. Claims: 52,72 completely; 1-44 partially
Idem as invention 1 but limited to SEQ ID Nos. 12 and 29.
13. Claims: 52,74 completely; 1-44 partially
Idem as invention 1 but limited to SEQ ID Nos. 13 and 30.
14. Claims: 52,75 completely; 1-44 partially
Idem as invention 1 but limited to SEQ ID Nos. 14 and 31.
15. Claims: 52,76 completely; 1-44 partially
Idem as invention 1 but limited to SEQ ID Nos. 15 and 32.
16. Claims: 52,77 completely; 1-44 partially
Idem as invention 1 but limited to SEQ ID Nos. 16 and 33.
17. Claims: 52,78 completely; 1-44 partially
Idem as invention 1 but limited to SEQ ID Nos. 17 and 34.

INTERNATIONAL SEARCH REPORT		International application No. PCT/US 01/19275
Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)		
This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:		
1.	<input checked="" type="checkbox"/> Claims Nos.:	
	because they relate to subject matter not required to be searched by this Authority, namely: see FURTHER INFORMATION sheet PCT/ISA/210	
2.	<input checked="" type="checkbox"/> Claims Nos.:	20, 21, 23, 24
	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: see FURTHER INFORMATION sheet PCT/ISA/210	
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(e).	
Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)		
This International Searching Authority found multiple inventions in this international application, as follows:		
see additional sheet		
1.	<input type="checkbox"/>	As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2.	<input type="checkbox"/>	As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3.	<input type="checkbox"/>	As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4.	<input checked="" type="checkbox"/>	No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
		1-44 partially; 45, 62 completely
Remark on Protest		
	<input type="checkbox"/>	The additional search fees were accompanied by the applicant's protest.
	<input type="checkbox"/>	No protest accompanied the payment of additional search fees.

International Application No. PCT/US 01/19275

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box 1.1

Although claim 18 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

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

Continuation of Box 1.2

Claims Nos.: 20,21,23,24

Present claims 20 and 23 relate to a composition comprising an agonist and antagonist, respectively, defined by reference to a desirable characteristic or property, namely being identified by the respective methods of claims 19 and 22. Furthermore, claims 21 and 24 relate to methods for treating a disease comprising administering said composition of claims 20 and 23, respectively.

The claims cover all agonists/antagonists having this characteristic or property, whereas the application provides no support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT for such agonists/antagonists.

In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT).

An attempt is made to define the product/compound/method/apparatus by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search over the whole of the claimed scope impossible. Consequently, no search has been carried out for claims 20,21,23,24.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter 11 procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.

PCT/US 01/19275

Patent document office in search report	Publication date	Patent family member(s)	Publication date	
WO 0002919	A	20-01-2000	EP WO 1097179 A1 0002919 A1	09-05-2001 20-01-2000
WO 0022129	A	20-04-2000	AU 6299199 A AU 6430799 A DE 1121431 T1 EP 1137776 A2 EP 1121431 A1 ES 2163384 T1 WO 0021967 A2 WO 0022129 A1 WO 0022131 A2	01-05-2000 01-05-2000 23-05-2002 04-10-2001 08-08-2001 01-02-2002 20-04-2000 20-04-2000 20-04-2000
WO 0015793	A	23-03-2000	AU 6035999 A EP 1114155 A2 WO 0015793 A2	03-04-2000 11-07-2001 23-03-2000
WO 9963087	A	09-12-1999	AU 4544999 A CA 2328891 A1 EP 1084241 A1 WO 9963087 A1	20-12-1999 09-12-1999 21-03-2001 09-12-1999
WO 0140797	A	07-06-2001	AU 1553101 A WO 0140797 A1 JP 2001309792 A	12-06-2001 07-06-2001 06-11-2001
WO 0144297	A	21-06-2001	AU 2095301 A WO 0144297 A1	25-06-2001 21-06-2001
WO 0175093	A	11-10-2001	US 2002061567 A1 AU 3486501 A AU 3995501 A AU 4925101 A AU 4972801 A AU 5119401 A AU 5521401 A WO 0157266 A1 WO 0164839 A2 WO 0175067 A2 WO 0175064 A2 WO 0174835 A1 WO 0175093 A1	23-05-2002 14-08-2001 12-09-2001 15-10-2001 15-10-2001 15-10-2001 15-10-2001 09-08-2001 07-09-2001 11-10-2001 11-10-2001 11-10-2001 11-10-2001

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 K 48/00	4 C 0 8 4
A 6 1 P 1/00	A 6 1 P 1/00	4 C 0 8 5
A 6 1 P 3/00	A 6 1 P 3/00	4 C 0 8 6
A 6 1 P 9/00	A 6 1 P 9/00	4 H 0 4 5
A 6 1 P 25/00	A 6 1 P 25/00	
A 6 1 P 31/00	A 6 1 P 31/00	
A 6 1 P 35/00	A 6 1 P 35/00	
A 6 1 P 37/00	A 6 1 P 37/00	
A 6 1 P 43/00	A 6 1 P 43/00	1 0 5
C 0 7 K 14/705	C 0 7 K 14/705	
C 1 2 N 1/15	C 1 2 N 1/15	
C 1 2 N 1/19	C 1 2 N 1/19	
C 1 2 N 1/21	C 1 2 N 1/21	
C 1 2 N 5/10	C 1 2 P 21/02	C
C 1 2 P 21/02	C 1 2 P 21/08	
C 1 2 P 21/08	C 1 2 Q 1/68	A
C 1 2 Q 1/68	G 0 1 N 33/15	Z
G 0 1 N 33/15	G 0 1 N 33/50	Z
G 0 1 N 33/50	G 0 1 N 33/53	D
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 60/216,595
(32)優先日 平成12年7月7日(2000.7.7)
(33)優先権主張国 米国(US)
(31)優先権主張番号 60/218,936
(32)優先日 平成12年7月14日(2000.7.14)
(33)優先権主張国 米国(US)
(31)優先権主張番号 60/219,154
(32)優先日 平成12年7月19日(2000.7.19)
(33)優先権主張国 米国(US)
(31)優先権主張番号 60/220,141
(32)優先日 平成12年7月21日(2000.7.21)
(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,CO,CR,CU,CZ,DE,DK,DM,DZ,EE,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NO,NZ,PL,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZW

- (72)発明者 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 ハファリア、エープリル・ジェイ・エイ
アメリカ合衆国カリフォルニア州9 5 0 5 4・サンタクララ・コーレデプリマベータ 2 2 2 7

- (72)発明者 ニュエン、ダニエル・ビー
アメリカ合衆国カリフォルニア州9 5 1 1 8・サンノゼ・リッジウッドドライブ 1 4 0 3
- (72)発明者 ガンディー、アミーナ・アール
アメリカ合衆国カリフォルニア州9 4 0 2 5・メンロパーク・# 1・ローブルアベニュー 8 3 7
- (72)発明者 カリック、デボラー・エイ
アメリカ合衆国カリフォルニア州9 4 0 2 5・メンロパーク・オリブストリート 9 0 0
- (72)発明者 グリフィン、ジェニファー・エイ
アメリカ合衆国カリフォルニア州9 4 5 5 5・フレモント・メローウェイ 3 3 6 9 1
- (72)発明者 ユエ、ヘンリー
アメリカ合衆国カリフォルニア州9 4 0 8 7・サニーベイル・ルイスアベニュー 8 2 6
- (72)発明者 カーン、ファラ・エイ
アメリカ合衆国カリフォルニア州9 4 0 4 0・マウンテンビュー・# 2 2 1・エスケラアベニュー
3 3 3
- (72)発明者 アービズ、チャンドラ・エス
アメリカ合衆国カリフォルニア州9 5 1 2 5・サンノゼ・モロッコドライブ 1 7 0 6
- (72)発明者 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 トリボレー、キャサリン・エム
アメリカ合衆国カリフォルニア州9 4 1 0 7・サンフランシスコ・# 5・テネシーストリート 1
1 2 1
- (72)発明者 リュ、ヤン
アメリカ合衆国カリフォルニア州9 4 3 0 3・パロアルト・コリーナウェイ 3 8 8 5
- (72)発明者 チョーラ、ナリンダー・ケイ
アメリカ合衆国カリフォルニア州9 4 5 8 7・ユニオンシティ・# 7 1 2・ユニオンスクエア 3
3
- (72)発明者 グラール、リチャード
アメリカ合衆国カリフォルニア州9 4 1 2 1・サンフランシスコ・トゥエンティナインスアベニ
ュー 6 8 2
- (72)発明者 ヤオ、モニック・ジー
アメリカ合衆国カリフォルニア州9 4 0 4 3・マウンテンビュー・フレデリックコート 1 1 1
- (72)発明者 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 ランクマール、ジャヤラクシミ
アメリカ合衆国カリフォルニア州9 4 5 5 5・フレモント・メイバードサークル 3 4 3 5 9
- (72)発明者 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・プリズベーン・ゴールデンイーグルレーン 2 3 3
- (72)発明者 エリオット、ビッキー・エス
アメリカ合衆国カリフォルニア州9 5 1 2 1・サンノゼ・ポルトンブレイスウェイ 3 7 7 0
- (72)発明者 ヘルナンデス、ロベルト
イギリス国ケント州・シーティー2 7 エスジー・カンタベリー・パインツリーアベニュー 1
6 0
- (72)発明者 ウォルシュ、ロドリック・ティー
イギリス国ケント州・シーティー1 3 イージー・カンタベリー・セントローレンスロード・バ
ウンダリーコート 8
- (72)発明者 ボロースキー、マーク・エル
アメリカ合衆国カリフォルニア州9 4 0 6 1・レッドウッドシティ・オーチャードアベニュー
1 2 2
- (72)発明者 ソーントン、マイケル
アメリカ合衆国カリフォルニア州9 4 0 6 2 - 2 6 1 2・ウッドサイド・メッドウェイロード 9
- (72)発明者 ヒー、アン

アメリカ合衆国カリフォルニア州 9 5 1 2 9 ・ サンノゼ ・ カタリナドライブ 4 6 0 1

F ターム(参考) 2G045 AA34 AA35 AA40 BA11 BB50 DA12 DA13 DA14 DA36 FB02
FB03
4B024 AA01 AA11 BA63 CA01 GA11 HA12 HA15
4B063 QA05 QA18 QA20 QQ02 QQ42 QQ53 QQ79 QQ96 QR08 QR32
QR42 QR55 QR62 QS25 QS33 QS34 QX01
4B064 AG01 AG27 CA19 CC24 DA01 DA13
4B065 AA93Y AB01 AC14 BA02 CA24 CA25 CA44 CA46
4C084 AA02 AA06 AA07 AA13 AA17 BA01 BA08 BA22 BA23 CA17
MA01 NA14 ZA012 ZA362 ZA662 ZB072 ZB112 ZB212 ZB262 ZB322
ZC212 ZC412
4C085 AA13 AA14 AA16 CC32 EE01 GG01
4C086 AA01 AA02 AA03 AA04 EA16 MA01 MA04 NA14 ZA01 ZA36
ZA66 ZB07 ZB11 ZB21 ZB26 ZB32 ZC21 ZC41
4H045 AA10 AA11 BA10 CA40 DA50 DA76 EA20 EA50 FA74

專利名稱(譯)	G蛋白偶聯受體		
公開(公告)號	JP2004523203A	公開(公告)日	2004-08-05
申請號	JP2002504306	申請日	2001-06-15
[標]申請(專利權)人(譯)	洞察Genomics公司		
申請(專利權)人(譯)	洞察基因組公司		
[標]發明人	ラルプリーティ ボーグンマライアアール ハファリアエープリルジェイエイ ニュエンダニエルビー ガンディーアミーナアール カリックデボラーエイ グリフィンジェニファーエイ ユエヘンリー カーンファラエイ アービズチャンドラエス リュデユングアイナエム トリボレーキャサリーンエム リュヤン チョーラナリンダーケイ グラールリチャード ヤオモニークジー ヤングジュンミング ランクマールジャヤラクシミ オウヤングジャンニス エリオットビッキーエス ヘルナンデスロベルト ウォルシュロドリックティー ボロースキーマークエル ソーントンマイケル ヒーアン		
發明人	ラル、プリーティ ボーグン、マライア・アール ハファリア、エープリル・ジェイ・エイ ニュエン、ダニエル・ビー ガンディー、アミーナ・アール カリック、デボラー・エイ グリフィン、ジェニファー・エイ ユエ、ヘンリー カーン、ファラ・エイ アービズ、チャンドラ・エス リュ、デユング・アイナ・エム トリボレー、キャサリーン・エム リュ、ヤン チョーラ、ナリンダー・ケイ グラール、リチャード ヤオ、モニーク・ジー ヤング、ジュンミング ランクマール、ジャヤラクシミ オウ・ヤング、ジャンニス		

エリオット、ビッキー・エス
ヘルナンデス、ロベルト
ウォルシュ、ロドリック・テイー
ボロースキー、マーク・エル
ソーントン、マイケル
ヒー、アン

IPC分类号	G01N33/50 A61K31/7088 A61K38/00 A61K39/395 A61K45/00 A61K48/00 A61P1/00 A61P3/00 A61P9/00 A61P25/00 A61P31/00 A61P35/00 A61P37/00 A61P43/00 C07K14/705 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N15/09 C12N15/12 C12P21/02 C12P21/08 C12Q1/68 G01N33/15 G01N33/53 G01N33/566
CPC分类号	A61K38/00 A61P1/00 A61P3/00 A61P9/00 A61P25/00 A61P31/00 A61P35/00 A61P37/00 A61P43/00 C07K14/705
FI分类号	C12N15/00.ZNA.A A61K31/7088 A61K39/395.D A61K39/395.N A61K45/00 A61K48/00 A61P1/00 A61P3/00 A61P9/00 A61P25/00 A61P31/00 A61P35/00 A61P37/00 A61P43/00.105 C07K14/705 C12N1/15 C12N1/19 C12N1/21 C12P21/02.C C12P21/08 C12Q1/68.A G01N33/15.Z G01N33/50.Z G01N33/53.D G01N33/53.M G01N33/566 C12N5/00.A A61K37/02
F-TERM分类号	2G045/AA34 2G045/AA35 2G045/AA40 2G045/BA11 2G045/BB50 2G045/DA12 2G045/DA13 2G045/DA14 2G045/DA36 2G045/FB02 2G045/FB03 4B024/AA01 4B024/AA11 4B024/BA63 4B024/CA01 4B024/GA11 4B024/HA12 4B024/HA15 4B063/QA05 4B063/QA18 4B063/QA20 4B063/QQ02 4B063/QQ42 4B063/QQ53 4B063/QQ79 4B063/QQ96 4B063/QR08 4B063/QR32 4B063/QR42 4B063/QR55 4B063/QR62 4B063/QS25 4B063/QS33 4B063/QS34 4B063/QX01 4B064/AG01 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/AA06 4C084/AA07 4C084/AA13 4C084/AA17 4C084/BA01 4C084/BA08 4C084/BA22 4C084/BA23 4C084/CA17 4C084/MA01 4C084/NA14 4C084/ZA012 4C084/ZA362 4C084/ZA662 4C084/ZB072 4C084/ZB112 4C084/ZB212 4C084/ZB262 4C084/ZB322 4C084/ZC212 4C084/ZC412 4C085/AA13 4C085/AA14 4C085/AA16 4C085/CC32 4C085/EE01 4C085/GG01 4C086/AA01 4C086/AA02 4C086/AA03 4C086/AA04 4C086/EA16 4C086/MA01 4C086/MA04 4C086/NA14 4C086/ZA01 4C086/ZA36 4C086/ZA66 4C086/ZB07 4C086/ZB11 4C086/ZB21 4C086/ZB26 4C086/ZB32 4C086/ZC21 4C086/ZC41 4H045/AA10 4H045/AA11 4H045/BA10 4H045/CA40 4H045/DA50 4H045/DA76 4H045/EA20 4H045/EA50 4H045/FA74
優先権	60/212483 2000-06-16 US 60/213954 2000-06-22 US 60/215209 2000-06-29 US 60/216595 2000-07-07 US 60/218936 2000-07-14 US 60/219154 2000-07-19 US 60/220141 2000-07-21 US
外部リンク	Espacenet

摘要(译)

本发明提供了鉴定和编码GCREC的人G蛋白偶联受体 (GCREC) 和多核苷酸。 本发明还提供表达载体， 宿主细胞， 抗体， 激动剂和拮抗剂。 本发明还提供了用于诊断， 治疗或预防与GCREC异常表达有关的疾病的方法。

logNo	航次号	logNo	航次号	logNo	航次号
752740	SSID No.	752740	SSID No.	752740	SSID No.
17468	1	17468	18	17468	17468
200743	2	200743	19	200743	200743
206305	3	206305	20	206305	206305
217432	4	217432	21	217432	217432
228292	5	228292	22	228292	228292
423211	6	423211	23	423211	423211
522832	7	522832	24	522832	522832
533317	8	533317	25	533317	533317
534348	9	534348	26	534348	534348
765264	10	765264	27	765264	765264
872412	11	872412	28	872412	872412
944476	12	944476	29	944476	944476
944476	13	944476	30	944476	944476
944476	14	944476	31	944476	944476
944476	15	944476	32	944476	944476
944476	16	944476	33	944476	944476
944476	17	944476	34	944476	944476