

(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2004-516817

(P2004-516817A)

(43) 公表日 平成16年6月10日(2004.6.10)

(51) Int. Cl. ⁷	F I	テーマコード (参考)
C 1 2 N 15/09	C 1 2 N 15/00	Z N A A 2 G O 4 5
A 6 1 K 38/00	A 6 1 K 39/395	D 4 B O 2 4
A 6 1 K 39/395	A 6 1 K 45/00	4 B O 6 3
A 6 1 K 45/00	A 6 1 P 1/04	4 B O 6 4
A 6 1 P 1/04	A 6 1 P 1/16	4 B O 6 5
	審査請求 未請求 予備審査請求 有	(全 256 頁) 最終頁に続く

(21) 出願番号	特願2002-516303 (P2002-516303)	(71) 出願人	301005050 インサイト・ゲノミックス・インコーポレイテッド アメリカ合衆国カリフォルニア州94304・パロアルト・ポータードライブ 3160
(86) (22) 出願日	平成13年7月25日 (2001.7.25)	(74) 代理人	100089266 弁理士 大島 陽一
(85) 翻訳文提出日	平成15年1月24日 (2003.1.24)	(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94062・ウッドサイド・メッドウェイロード 9
(86) 国際出願番号	PCT/US2001/023433		
(87) 国際公開番号	W02002/010387		
(87) 国際公開日	平成14年2月7日 (2002.2.7)		
(31) 優先権主張番号	60/221, 478		
(32) 優先日	平成12年7月27日 (2000.7.27)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/223, 268		
(32) 優先日	平成12年8月3日 (2000.8.3)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/227, 054		
(32) 優先日	平成12年8月21日 (2000.8.21)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 Gタンパク質共役受容体

(57) 【要約】

本発明はヒトGタンパク質共役受容体 (GCRRC)、およびGCRRC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、GCRRC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 - 19 (配列番号 1 乃至 19) 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 - 19 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アミノ酸配列を含む天然のポリペプチド

(c) SEQ ID NO: 1 - 19 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 - 19 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原性断片 10

【請求項 2】

SEQ ID NO: 1 - 19 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: 20 - 38 (配列番号 20 乃至 38) 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リンク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リンク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: 20 - 38 からなる群から選択したポリヌクレオチド配列を含むポリヌクレオチド 40

(b) SEQ ID NO: 20 - 38 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からなり、前記プローブと前記標的ポリヌクレオチドまたは断片の間でハイブリダイゼーション複合体が形成されるような条件下で、前記プローブが前記標的ポリヌクレオチドに特異的にハイブリダイズすることを特徴とする方法。

10

【請求項 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。

【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。

20

【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。

【請求項 1 7】

前記ポリペプチドが、SEQ ID NO : 1 - 1 9 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。

30

【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、

(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。

40

【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。

【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。

50

【請求項 23】

請求項 22 に記載の方法によって同定したアンタゴニスト化合物と、薬剤として許容できる賦形剤とを含む組成物。

【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、

10

(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。

【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、

20

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

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

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

【請求項 27】

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

30

(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。

【請求項 28】

試験化合物の毒性を算定する方法であって、

(a) 核酸を含む生物学的サンプルを前記試験化合物で処理する過程と、

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

40

(c) ハイブリダイゼーション複合体の収量を定量するステップと、

(d) 前記処理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、

50

(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。

10

【請求項31】

請求項10に記載の抗体と、許容できる賦形剤とを含む化合物。

【請求項32】

被検者の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。

20

【請求項35】

請求項10に記載の抗体の特異性を有するポリクローナル抗体を調製する方法であって、

(a) 抗体反応を誘発する条件下で、SEQ ID NO: 1-19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: 1-19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。

30

【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: 1-19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: 1-19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。

40

【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】

50

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】

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

S E Q I D N O : 1 - 9 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、

10

(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、S E Q I D N O : 1 - 19 を有する群から選択したアミノ酸配列を有するポリペプチドがサンプル中に存在することを標示することを特徴とする方法。

【請求項 44】

S E Q I D N O : 1 - 19 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、

20

(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、S E Q I D N O : 1 - 19 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。

【請求項 45】

S E Q I D N O : 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】

S E Q I D N O : 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】

S E Q I D N O : 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】

S E Q I D N O : 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。

30

【請求項 49】

S E Q I D N O : 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。

【請求項 50】

S E Q I D N O : 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】

S E Q I D N O : 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】

S E Q I D N O : 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】

S E Q I D N O : 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。

40

【請求項 54】

S E Q I D N O : 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。

【請求項 55】

S E Q I D N O : 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】

S E Q I D N O : 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】

S E Q I D N O : 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】

50

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。

【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 63】

SEQ ID NO : 19 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 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。

10

20

30

40

50

【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。

【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。

【請求項 79】

SEQ ID NO : 35 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 80】

SEQ ID NO : 36 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 81】

SEQ ID NO : 37 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 82】

SEQ ID NO : 38 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タンパク質共役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タンパク質共役受容体の核酸配列及びアミノ酸配列の発現における外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。

【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される。このように活性化された受容体が細胞内の生化学的カスケードを誘発し、このカスケードは転写因子などの細胞内標的分子の活性化で終る。シグナル伝達のこのプロセスは、細胞増殖、分化及び遺伝子転写を含む全てのタイプの細胞機能を制御する。同定されている遺伝子の最も大きいファミリーの1つによってコードされるGタンパク質共役受容体(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。

【0003】

GPCRは、7つの疎水性膜貫通ドメインを有することを特徴とする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()螺旋の束を形成している。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、膜貫通ドメインに連結している。第2及び第3細胞外ループを結びつけ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。膜貫通ドメインによって、受容体の構造的及び機能的特徴がある程度決まる。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

10

20

30

40

50

成する。細胞外 N 末端セグメント、または 3 つの細胞外ループの内の 1 つまたは複数のループが、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 の細胞内ループがヘテロ 3 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し、この G タンパク質複合体が更に細胞内のシグナル伝達作用を仲介する。この細胞内のシグナル伝達作用には、サイクリック AMP (cAMP)、ホスホリパーゼ C、およびイノシトール三リン酸などのセカンドメッセンジャーの活性化、並びに活性化された GPCR とイオンチャネルタンパク質との相互作用が含まれる (Watson, S. および S. Arkininstall (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego CA, 2-6 頁; Bolander, F. F. (1994) Molecular Endocrinology, Academic Press, San Diego CA, 162-176 頁; Baldwin, J. M. (1994) Curr. Opin. Cell Biol. 6: 180-190 を参照)。

10

【0004】

GPCR には、感覚性シグナルメディエータの受容体 (例えば光及び嗅覚刺激性分子)、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。刺激がまだ同定されていない受容体として作用する GPCR は、オーファン受容体として知られている。

20

【0005】

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ている受容体は現在では 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。N 末端及び細胞質ループの変異体もまた高頻度であるのに対して、細胞外ループまたは膜貫通ドメインでの変異体は頻度が低い。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)。

30

【0006】

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、神経伝達物質への応答に直接関与する。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

40

50

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頁, 19-22頁, 32-35頁, 130-131頁, 214-216頁, 221-222頁、Habert-Ortoli, E. ら (1994) Proc. Natl. Acad. Sci. USA 91:9780-9783に概説されている)。

【0007】

ガラニン受容体は、インスリン、アセチルコリン、セロトニン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、摂食障害、痛み、うつ病及びアルツハイマー病に關与している (Kask, K. ら (1997) Life Sci. 60:1523-1533)。その他の神経系ロドプシン様GPCRには、発達及び神経病理学において役割を持つと考えられるリゾホスファチジン酸、およびその他のリゾリン脂質に対する受容体など、増大する数のファミリーが含まれる (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、嗅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、多数の異なる嗅覚受容体が必要である。各嗅覚知覚ニューロンは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、脳の非常に特有の領域及び嗅覚上皮の特定の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) 受容体 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)。

【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頁)。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。

【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モジュール含有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. および S. Gordon (1998) J. Leukoc. Biol. 63:271-280)。

【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、長期の増強作用に關与する (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

10

20

30

40

50

を検出するものであり、カルシウム結合に關与し得る酸性アミノ酸のクラスターを含む大きな細胞外ドメインを有する。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。

【0012】

その他のGPCRのサブファミリーには、線虫(*Caenorhabditis elegans*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遠く関係している。細胞膜上の接合因子への反応に關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(*Dictyostelium discideum*)で個々の細胞の集合を調整し、多く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。

10

【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、多数のヒト疾患に關係している(前出のCoughlin)。例えば、色素性網膜炎はロドプシン遺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 -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、ヒトの疾患に關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關与している(Horn, F. and G. Vriend(1998) *J. Mol. Med.* 76: 464-468)。

20

【0014】

更に、過去20年間にGPCRの活性化及び阻害を誘発する数百の新薬が認められてきている。これらの薬剤の治療標的は、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ハンチントン病の初期段階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)。

30

40

【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手段の可能性を示唆する(Schoneberg, T. ら(1996) *EMBO J.* 15: 1283-1291)。メラノコルチン4受容体(MC4R)における突然変異は、ヒトの体重調節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と同様に、これらMC4Rの突然変異体は原形質膜への輸送中に欠陥があるので(Ho, G. and R.G. MacKenzie(1999) *J. Biol. Chem.* 27

50

4 : 3 5 8 1 6 - 3 5 8 2 2)、同様の手段で治療し得る。副甲状腺ホルモン (P T H) 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 P T H 依存型制御を仲介する G P C R である。P T H / 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 P T H 受容体リガンドの開発を可能にし得る (M a n n s t a d t , M . ら (1 9 9 9) A m . J . P h y s i o l . 2 7 7 : F 6 6 5 - F 6 7 5)。

【 0 0 1 6 】

G P C R 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 (レビューは、L o c a t i , M . a n d P . M . M u r p h y (1 9 9 9) A n n u . R e v . M e d . 5 0 : 4 2 5 - 4 4 0 を参照のこと)。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は、病的炎症において及び多発性硬化症などの自己免疫疾患においてこれらの受容体が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 (H I V - 1) を含む病原菌、によっても、ケモカイン受容体は感染を促進するために利用される。ケモカイン受容体 C C R 5 の短縮変種は、H I V - 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、A I D S に対する抵抗力を生じさせ、C C R 5 のアンタゴニストが A I D S の発達予防に有用たり得ることを示唆する。

10

【 0 0 1 7 】

新たな G タンパク質共役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性感症の診断、治療並びに予防と、G タンパク質共役受容体の核酸配列及びアミノ酸配列の発現における外因性化合物の効果の算定において有用である新たな組成を提供することにより、当分野における要求を満たす。

20

【 0 0 1 8 】

(発明の概要)

本発明は、総称して「 G C R E C 」、個別にはそれぞれ「 G C R E C - 1 」、 「 G C R E C - 2 」、 「 G C R E C - 3 」、 「 G C R E C - 4 」、 「 G C R E C - 5 」、 「 G C R E C - 6 」、 「 G C R E C - 7 」、 「 G C R E C - 8 」、 「 G C R E C - 9 」、 「 G C R E C - 1 0 」、 「 G C R E C - 1 1 」、 「 G C R E C - 1 2 」、 「 G C R E C - 1 3 」、 「 G C R E C - 1 4 」、 「 G C R E C - 1 5 」、 「 G C R E C - 1 6 」、および「 G C R E C - 1 7 」と呼ぶ G タンパク質共役受容体である精製されたポリペプチドを提供する。或る実施態様において本発明は、(a) S E Q I D N O : 1 - 1 9 を有する群から選択したアミノ酸配列からなるポリペプチド、(b) S E Q I D N O : 1 - 1 9 を有する群から選択したアミノ酸配列と少なくとも 9 0 % 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) S E Q I D N O : 1 - 1 9 を有する群から選択したアミノ酸配列を有するポリペプチドの生物学的活性断片、または(d) S E Q I D N O : 1 - 1 9 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原性断片を含む群から選択した実質上単離されたポリペプチドを提供する。一実施態様では、S E Q I D N O : 1 - 1 9 のアミノ酸配列を含む実質上単離されたポリペプチドを提供する。

30

【 0 0 1 9 】

また、本発明は(a) S E Q I D N O : 1 - 1 9 を有する群から選択したアミノ酸配列からなるポリペプチド、(b) S E Q I D N O : 1 - 1 9 を有する群から選択したアミノ酸配列と少なくとも 9 0 % 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) S E Q I D N O : 1 - 1 9 を有する群から選択したアミノ酸配列を有するポリペプチドの生物学的活性断片、または(d) S E Q I D N O : 1 - 1 9 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 S E Q I D N O : 1 - 1 9 を有する群から選択したポリペプチドをコードする。別の実施態様では、ポリヌクレオチドは S E Q I D N O : 2 0 - 3 8 を有する群から選択される。

40

50

【0020】

本発明は更に、(a) SEQ ID NO: 1 - 19 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 - 19 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 - 19 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 - 19 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リンク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。 10

【0021】

更に本発明は、(a) SEQ ID NO: 1 - 19 からなる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 - 19 からなる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 - 19 からなる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 - 19 とからなる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を製造する方法を提供する。製造方法は、(a) 組換えポリヌクレオチドを用いて形質転換した細胞をポリペプチドの発現に適した条件下で培養する過程と、(b) そのように発現したポリペプチドを受容する過程とを有し、組換えポリヌクレオチドはポリペプチドをコードするポリヌクレオチドに機能的にリンクしたプロモーター配列を有する。 20

【0022】

本発明は更に、(a) SEQ ID NO: 1 - 19 からなる群から選択したアミノ酸配列を有するポリペプチド、(b) SEQ ID NO: 1 - 19 からな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含むポリペプチド、(c) SEQ ID NO: 1 - 19 からなる群から選択したアミノ酸配列を有するポリペプチドの生物学的活性断片、および(d) SEQ ID NO: 1 - 19 からなる群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片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。 30

【0023】

本発明は更に、(a) SEQ ID NO: 20 - 38 からなる群から選択したポリヌクレオチド配列、(b) SEQ ID NO: 20 - 38 からなる群から選択したポリヌクレオチド配列と少なくとも90%の同一性を有する天然のポリヌクレオチド配列を含むポリヌクレオチド、(c) (a) に相補的なポリヌクレオチド配列、(d) (b) に相補的なポリヌクレオチド配列、および(e) (a) ~ (d) のRNA等価物からなる群から選択された単離されたポリヌクレオチドを提供する。一実施態様では、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。 40

【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する。ここで、標的ポリヌクレオチドは(a) SEQ ID NO: 20 - 38 からなる群から選択したポリヌクレオチド配列を含むポリヌクレオチド、(b) SEQ ID NO: 20 - 38 からなる群から選択したポリヌクレオチド配列と少なくとも90%の同一性を有する天然のポリヌクレオチド配列を含むポリヌクレオチド、(c) (a) に相補的なポリヌクレオチド、(d) (b) に相補的なポリヌクレオチド配列、および(e) (a) ~ (d) のRNA等価物からなる群から選択されたポリヌクレオチド配列を含む。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を含む少なくとも20の連続したヌクレオチドを含むプローブを用いて該サンプルをハイブリダイズする過程と、(b) ハイブリダイゼーション複合体の存在・不存在を検出し、複合体が存在する場合にはオプションでそ 50

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: 20-38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: 20-38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。

10

【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: 1-19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: 1-19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: 1-19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: 1-19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原性断片を含む。一実施例では、SEQ ID NO: 1-19からなる群から選択されたアミノ酸配列を含む組成物を提供する。更に、本発明は、患者にこの組成物を投与することを含む、機能的GCRECの発現の低下に関連した疾患やその症状の治療方法を提供する。

20

【0027】

本発明はまた、(a)SEQ ID NO: 1-19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: 1-19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: 1-19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: 1-19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GCREC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。

30

40

【0028】

本発明は更に、(a)SEQ ID NO: 1-19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: 1-19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: 1-19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: 1-19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)

50

サンプル中のアゴニスト活性を検出する過程とを含む。一実施態様で本発明は、この方法によって同定したアンタゴニスト化合物と薬剤として許容できる賦形剤とを含む成分を提供する。別の実施態様では、機能性GCREC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。

【0029】

本発明は更に、(a)SEQ ID NO: 1-19からな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: 1-19からなる群から選択したアミノ酸配列と少なくとも90%の同一性を有する天然のアミノ酸配列を含むポリペプチド、(c)SEQ ID NO: 1-19を有する群から選択したアミノ酸配列の生物学的活性断片、または(d)SEQ ID NO: 1-19からなる群から選択したアミノ酸配列の免疫原性断片を含む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。

10

【0030】

本発明は更に、(a)SEQ ID NO: 1-19を有する群から選択したアミノ酸配列、(b)SEQ ID NO: 1-19を有する群から選択したアミノ酸配列と少なくとも90%の相同性を有する天然のアミノ酸配列、(c)SEQ ID NO: 1-19を有する群から選択したアミノ酸配列の生物学的活性断片、または(d)SEQ ID NO: 1-19を有する群から選択したアミノ酸配列の免疫原性断片を含むポリペプチドの活性を調節する化合物をスクリーニングする方法を提供する。スクリーニング方法は、(a)ポリペプチドの活性が許容された条件下で、ポリペプチドを少なくとも1つの試験化合物と混合させる過程と、(b)ポリペプチドの活性を試験化合物の存在下で算定する過程と、(c)試験化合物の存在下でのポリペプチドの活性を試験化合物の不存在下でのポリペプチドの活性と比較する過程とを含み、試験化合物の存在下でのポリペプチドの活性の変化は、ポリペプチドの活性を調節する化合物であることを意味する。

20

【0031】

更に本発明は、SEQ ID NO: 20-38からなる群から選択された配列を含む標的ポリヌクレオチドの発現を変化させるのに効果的な化合物をスクリーニングする方法であって、(a)この標的ポリヌクレオチドを含むサンプルを化合物に曝露するステップと、(b)この標的ポリヌクレオチドの発現の変化を検出するステップとを含む、該スクリーニング方法を提供する。

30

【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: 20-38からなる群から選択したポリヌクレオチド配列を含むポリヌクレオチド、(ii)SEQ ID NO: 20-38からなる群から選択したポリヌクレオチド配列と少なくとも90%の同一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等価物からなる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を含むプローブを用いて、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: 20-38からなる群から選択したポリヌクレオチド配列を含むポリヌクレオチド、(ii)SEQ ID NO: 20-38からなる群から選択したポリヌクレオチド配列と少なくとも90%の同一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、お

40

50

よび (v) (i) ~ (i v) の R N A 等価物からなる群から選択する。或いは、標的ポリヌクレオチドは、上記 (i) ~ (v) からなる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。

【 0 0 3 3 】

(本発明の記載について)

本発明のタンパク質、ヌクレオチド配列及び方法について説明するが、その前に、説明した特定の装置、材料及び方法に本発明が限定されるものではなく、改変し得ることを理解されたい。また、ここで使用する専門用語は特定の実施例を説明する目的で用いたものに過ぎず、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。

10

【 0 0 3 4 】

請求の範囲及び明細書中で用いている単数形の「或る」及び「その(この)」の表記は、文脈から明らかにそうでないとされる場合を除いて複数のも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。

【 0 0 3 5 】

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

20

【 0 0 3 6 】

(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。

30

【 0 0 3 7 】

用語「アゴニスト」は、G C R E C の生物学的活性を強めたり、模倣する分子を指す。アゴニストの例として、タンパク質、核酸、糖質、小分子その他の任意の化合物や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エレメントに作用することによって、G C R E C の活性を調節する。

【 0 0 3 8 】

用語「対立遺伝子変異配列」は、G C R E C をコードする遺伝子の別の形を指す。対立遺伝子変異体は、核酸配列における少なくとも 1 の突然変異から作製し得る。また、変異 R N A 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。

40

【 0 0 3 9 】

G C R E C をコードする「変異」核酸配列は、様々なヌクレオチドの欠失、挿入、或いは

50

置換が起こっても、G C R E Cと同じポリペプチド或いはG C R E Cの機能特性の少なくとも1つを備えるポリペプチドを指す。この定義に含まれるのは、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予期しない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され得、サイレント変化を生じG C R E Cと機能的に等価となるアミノ酸残基の欠失、挿入、或いは置換を含み得る。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基づき行い得る。例えば、負に帯電したアミノ酸にはアスパラギン酸及びグルタミン酸があり、正に帯電したアミノ酸にはリジン及びアルギニンがある。親水性値が近似している非荷電極性側鎖を有するアミノ酸には、アスパラギンとグルタミン、セリンとスレオニンがある。親水性値が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。

10

20

30

40

50

【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。

【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】

用語「アンタゴニスト」は、G C R E Cの生物学的活性を阻害或いは減弱する分子である。アンタゴニストとしては、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エレメントに作用することによって、G C R E Cの活性を調節する。

【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。G C R E Cポリペプチドを結合する抗体は、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シガイのヘモシアニン（K L H）等がある。結合ペプチドは、動物を免疫化するために用いる。

【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、ホスホロチオ酸、メチルホスホン酸またはベンジルホスホン酸

等の修飾されたバックボーン連鎖を有するオリゴヌクレオチドや、2'-メトキシエチル糖または2'-メトキシエトキシ糖等の修飾された糖類を有するオリゴヌクレオチドや、或いは5-メチルシトシン、2-デオキシウラシルまたは7-デアザ-2'-デオキシグアノシン等の修飾された塩基を有するオリゴヌクレオチドが含まれる。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」若しくは「マイナス(-)」の語が参照DNA分子のアンチセンス鎖を、「正」若しくは「プラス(+)」がセンス鎖を指すことがある。

【0046】

10

「生物学的に活性」の語は、天然分子の構造的機能、調節機能または生化学的機能を有するタンパク質を指す。同様に、用語「免疫学的に活性」または「免疫原性」は、天然或いは組換え体のGCRC、合成のGCRCまたはそれらの任意のオリゴペプチドが、適当な動物或いは細胞の特定の免疫応答を誘発して特定の抗体と結合する能力を指す。

【0047】

「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'」に結合する。

【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。GCRC若しくはGCRCの断片をコードするポリヌクレオチド配列を含む組成物は、ハイブリダイゼーションプローブとして使用され得る。このプローブは、凍結乾燥状態で保存可能であり、糖質などの安定化剤と結合させることが可能である。ハイブリダイゼーションにおいては、塩(例えばNaCl)、界面活性剤(例えばドデシル硫酸ナトリウム; SDS)及びその他の構成エレメント(例えばデンハート液、脱脂粉乳、サケの精子のDNA等)を含む水溶液中にプローブを分散させることができる。

20

【0049】

「コンセンサス配列」は、不要な塩基を分離するためにDNA配列の解析を繰り返し行い、XL-PCRキット(PE Biosystems, Foster City CA)を用いて5'及び/または3'の方向に伸長され、再度シーケンシングされた核酸配列、またはGELVIEW断片構築システム(GCG, Madison, WI)またはPhrap(University of Washington, Seattle WA)等の断片構築用のコンピュータプログラムを用いて1つ或いはそれ以上の重複するcDNAやEST、またはゲノムDNA断片から構築された核酸配列を指す。伸長及び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。

30

【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い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。

40

元残基	保的置換	
Aa	Gly, Ser	
Ag	His, Lys	
An	Asp, Glu, His	
Ap	Asn, Glu	
Gs	Ala, Ser	10
Gh	Asn, Glu, His	
Gu	Asp, Glu, His	
Gly	Ala	
His	Asn, Arg, Glu, Gu	
Ile	Ileu, Val	
Ileu	Ile, Val	20
Lys	Arg, Glu, Gu	
Met	Ileu, Ile	
Phe	His, Met, Ileu, Trp, Tyr	
Ser	Gly, Thr	
Thr	Ser, Val	
Trp	Phe, Tyr	30
Tyr	His, Phe, Trp	
Val	Ile, Ileu, Thr	

【 0 0 5 1 】

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。 40

【 0 0 5 2 】

「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。

【 0 0 5 3 】

「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 (p e g y l a t i o n) 50

)、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。

【0054】

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

【0055】

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

10

【0056】

「エキソンシャフリング」は異なったコード領域(エキソン)の組換えを指す。エキソンはコードされたタンパク質の構造的、または機能的ドメインを示すので、新規タンパク質は安定な下部構造の新しい再組合せによって組み立てられ、従って新しいタンパク質機能の発展が促進され得る。

【0057】

用語「断片」は、G C R E CまたはG C R E Cをコードするポリヌクレオチドの固有の部分であって、その親配列(parent sequence)と同一であるがその配列より長さが短いものを指す。断片は、画定された配列の全長から1ヌクレオチド/アミノ酸残基を差し引いた長さよりも短い長さを有し得る。例えば或る断片は、5~1000の連続したヌクレオチドまたはアミノ酸残基を有し得る。プローブ、プライマー、抗原、治療用分子として、或いはその他の目的のために用いられる断片は、少なくとも5、10、15、20、25、30、40、50、60、75、100、150、250若しくは500の連続したヌクレオチド或いはアミノ酸残基長さであり得る。断片は、分子の特定領域から優先的に選択し得る。例えば、ポリペプチド断片は、所定の配列に示すような最初の250または500アミノ酸(またはポリペプチドの最初の25%または50%)から選択された或る長さの連続したアミノ酸を有し得る。これらの長さは明らかに例として挙げているものであり、本発明の実施例では、配列表、表及び図面を含む明細書に裏付けされた任意の長さであってよい。

20

30

【0058】

SEQ ID NO: 20 - 38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: 20 - 38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: 20 - 38以外の配列とは異なるものである。SEQ ID NO: 20 - 38のある断片は、例えば、ハイブリダイゼーションや増幅技術、またはSEQ ID NO: 20 - 38を関連ポリヌクレオチド配列から区別する類似の方法に有用である。SEQ ID NO: 20 - 38の断片の正確な長さ及び断片に対応するSEQ ID NO: 20 - 38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。

【0059】

SEQ ID NO: 1 - 19の断片は、SEQ ID NO: 20 - 38の断片によってコードされる。SEQ ID NO: 1 - 19の断片には、SEQ ID NO: 1 - 19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 - 19の断片は、SEQ ID NO: 1 - 19を特異認識する抗体を産出するための免疫原性ペプチドとして有用である。SEQ ID NO: 1 - 19の断片及び断片に対応するSEQ ID NO: 1 - 19の断片の正確な長さとその断片に対応するSEQ ID NO: 1 - 19での領域は、その断片の目的に基づいて当分野で一般的な技術によって日常的に決定できる。

40

【0060】

「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

50

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含む配列である。「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。

【0061】

「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。

【0062】

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。

10

【0063】

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

20

【0064】

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

30

40

Matrix: HCSUM2

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x d r p d f: 50

Expect: 10

Word Size: 11

Filter: on

10

【0065】

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。

20

【0066】

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。

【0067】

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。

30

【0068】

ポリペプチド配列間の一致率は、MEGALIGN version 3.12e 配列アラインメントプログラムに組み込まれているようなCLUSTAL Vアルゴリズムのデフォルトのパラメータを用いて決定できる（既に説明したのでそれを参照されたい）。CLUSTAL Vを用いる対方式のポリペプチド配列のアラインメントの場合、デフォルトパラメータは、K t u p l e = 1、g a p p e n a l t y = 3、w i n d o w = 5、及び「d i a g o n a l s s a v e d」= 5と設定する。PAM250マトリクスが、デフォルトの残基重み付け表として選択される。ポリヌクレオチドアラインメントと同様に、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。

40

【0069】

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

50

示す。

【 0 0 7 0 】

Matrix: HCGUM2

OpenCap: 11 and ExtensionCap: 1 penalties

Gap: d: pdf: 50

Expect: 10

WordSize: 3

Filter: on

10

【 0 0 7 1 】

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。

20

【 0 0 7 2 】

「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エレメントを含む直鎖状の小染色体である。

【 0 0 7 3 】

「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。

30

【 0 0 7 4 】

「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。

40

【 0 0 7 5 】

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

50

ヨ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。

【0076】

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 x SSC及び約0.1%のSDSの存在下、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 x SSCの範囲で変化し得る。通常は、ブロッキング剤を用いて非特異ハイブリダイゼーションを阻止する。このような遮断剤には、例えば、約100~200 µg/mlの変性サケ精子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

【0077】

「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。

20

【0078】

「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。

【0079】

「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。

30

【0080】

「免疫原性断片」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。

【0081】

「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。

40

【0082】

「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。

【0083】

用語「調節」は、GCRCの活性の変化を指す。調節することによって例えば、GCRC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。

【0084】

「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

50

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。

【0085】

「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

【0086】

「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。

【0087】

GCRECの「翻訳後修飾」には、脂質化、グリコシル化、リン酸化、アセチル化、ラセミ化、蛋白分解性切断及びその他の当分野で既知の修飾が含まれ得る。これらのプロセスは、合成的或いは生化学的に生じ得る。生化学的修飾は、GCRECの酵素環境に依存し、細胞タイプによって変化し得る。

20

【0088】

「プローブ」とは、同一配列或いは対立遺伝子核酸配列、関連する核酸配列の検出に用いる、GCRECやそれらの相補配列、またはそれらの断片をコードする核酸配列のことである。プローブは、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鎖に伸長し得る。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

【0089】

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

【0090】

プローブ及びプライマーの調製及び使用方法については、Sambrook, J. ら(1989) Molecular Cloning: A Laboratory Manual, 第2版, 1-3巻, Cold Spring Harbor Press, Plainview NY、Ausubel, F.M. ら, (1987) Current Protocols in Molecular Biology, Greene Publi. Assoc. & Wiley-Intersciences, New York NY、Innis ら(1990) PCR Protocols, A Guide to Methods and Applications Academic Press, S

50

an 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。

【0091】

プライマーとして用いるオリゴヌクレオチドの選択は、そのような目的のために本技術分野でよく知られているソフトウェアを用いて行う。例えばOLIGO 4.06ソフトウェアは、各100ヌクレオチドまでのPCRプライマー対の選択に有用であり、オリゴヌクレオチド及び最大5,000までの大きめのポリヌクレオチドであって32キロベースまでのインプットポリヌクレオチド配列から得たものを分析するのにも有用である。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。

【0092】

「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。

【0093】

或いはこのような組換え核酸は、ウイルスベクターの不可欠なエレメントであって例えばワクシニアウイルスに基づくものであり得る。ワクシニアウイルスは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。

【0094】

「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。

10

20

30

40

50

【0095】

「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。

【0096】

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。

【0097】

「サンプル」の語は、その最も広い意味で用いられる。GCRECをコードする核酸若しくはその断片、またはGCREC自体を含む疑いのあるサンプルは、体液や、細胞から単離された細胞、染色体、細胞小器官または膜からの抽出物や、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。

【0098】

「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

10

20

【0099】

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

【0100】

「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。

30

【0101】

「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。

【0102】

「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。

40

【0103】

「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（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

50

含まれる。

【0104】

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

10

【0105】

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「多型性」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。多型性変異体は、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、多型性変異体は、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

20

30

【0106】

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

【0107】

(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の発見と、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、治療並びに予防におけるこれらの組成物の利用に基づくものである。

【0108】

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Incycyteプロジェクト識別番号(IncycyteプロジェクトID)と関連する。各ポリペプチド配列は、ポリ

50

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。

【0109】

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

【0110】

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、Gwn Bank 相同体のアノテーションを示す。

【0111】

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。

【0112】

合わせて、表2及び表3は本発明の各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タンパク質共役受容体であることを証明する。例えば、SEQ ID NO: 1は Meleagris gallopavo Gタンパク質共役受容体、P2Yヌクレオチド受容体 (GenBank ID g2707256) と40%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 $4e^{-62}$ 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、ロドプシンファミリーの膜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 および BLAST 分析から得たデータは、SEQ ID NO: 1がGタンパク質共役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は、同様の規則で解析され注釈された。これらの分析は、SEQ ID NO: 2はフェロモン受容体であることを示す (Dulac, C. 及び R. Axel (1995) Cell 83: 195-206)。

【0113】

別の例において、SEQ ID NO: 6は Basic Local Alignment Search Tool (BLAST) によって同定されたヒトC-Cケモカイン受容体1型 (GenBank ID g179985) に29%同一である。(表2参照) BLAST 確率スコアは $1.6e^{-15}$ であり、これは観察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: 6はまた、保存されたタンパク質ファミリードメインの隠れマルコフモデル (HMM) ベースPFAMデータベ

ースにおいて統計学的に有意な一致を検索することにより決定されるような膜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(表3を参照)。BLIMPS及びPROFILES CAN解析よりのデータは、SEQ ID NO:6がケモカイン受容体である、さらに実証的な証拠を提供する。

【0114】

例えば、SEQ ID NO:9はラットカルシウム非依存性 - ラトロトキシン受容体(GenBank ID g3882981)と95%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0.0であるが、これは観測されたポリペプチド配列が偶然に得られる確率を示す。SEQ ID NO:9はまた、膜7回貫通受容体(セクレチンファミリー)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PROFILES CAN分析から得たデータは、SEQ ID NO:9がラトロフィリン関連Gタンパク質共役受容体であることを裏づける証拠を更に提供する。

10

【0115】

例えば、SEQ ID NO:12はマウスGタンパク質共役受容体、GPR73(GenBank ID g7248884)と84%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 $6.7e-166$ であり、これは観測されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO:12はまた、膜7回貫通(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分析から得られたデータによって、ロドプシン様GPCRスーパーファミリーのシグネチャの存在が明らかにされる。MOTIFS およびPROFILES CAN 分析から得たデータは、SEQ ID NO:12がGタンパク質共役受容体であることを裏づける証拠を更に提供する。

20

【0116】

例えば、SEQ ID NO:15はラットセロトニン受容体(GenBank ID g310075)と80%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 $2.5e-152$ 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:15はまた、ロドプシンファミリー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分析から得たデータは、SEQ ID NO:15がGタンパク質共役受容体であることを裏づける証拠を更に提供する。

30

【0117】

例えば、SEQ ID NO:16はマウス嗅覚受容体E3(GenBank ID g3983382)と71%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 $1.9e-88$ 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:16はまた、ロドプシンファミリーの膜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PROFILES CAN分析から得たデータは、SEQ ID NO:16が嗅覚Gタンパク質共役受容体であることを裏づける証拠を更に提供する。例えば、SEQ ID N

40

50

O : 17 はマウス嗅覚 G タンパク共役受容体 G3 (GenBank ID g3983398) と 83 % 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 $5.0e-99$ であり、これは観測されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO : 17 はまた、ロドプシンファミリーの膜 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 PROFILSCAN 分析から得たデータは、SEQ ID NO : 17 が嗅覚 G タンパク質共役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 - 5、SEQ ID NO : 7 - 8、SEQ ID NO : 10 - 11、SEQ ID NO : 13 - 14 および SEQ ID NO : 18 - 19 は同様にして解析し、注釈付けをされた。SEQ ID NO : 1 - 19 の解析のためのアルゴリズム及びパラメータは表7に記述する。

【0118】

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る構築した。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 Polynucleotide 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 : 20 - 38 を同定するため、或いは SEQ ID NO : 20 - 38 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構築するのに用いた。表4の列6および列7はそれぞれ、列5の配列に対応するcDNA 配列および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。

【0119】

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、7075196H1 は Incyte cDNA 配列の識別番号であり、BRAUTDR04 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、71906055V1) に由来する。または、列5の識別番号は、ポリヌクレオチド配列の組み立てに用いた GenBank の cDNA すなわち EST (例えば、g900324) の識別番号の場合もある。さらに、列5の識別番号は ENSEMBL (The Sanger Centre, Cambridge, UK) データベース (すなわち、「ENST」と命名された配列) から由来した配列を示す。あるいは、列5の識別番号は NCBI RefSeq ヌクレオチド配列記録データベース (すなわち、「NM」または「NT」と命名された配列) または NCBI RefSeq タンパク質配列記録 (すなわち、「NP」と命名された配列) から由来している。または列5の識別番号は、「エキソンスティック (exon - stitching)」アルゴリズムにより結び合わせた cDNA 及び Genscan 予想エキソンの両方の集合を指す場合もある。例えば、FL _ X X X X X X _ N₁ _ N₂ _ Y Y Y Y Y _ N₃ _ N₄ は「スティッチ」配列を表しており、X X X X X X はアルゴリズムが適用された配列のクラスターの識別番号であり、また、Y Y Y Y Y はアルゴリズムによって出された予測の番号である。N₁, N₂, N₃, ... が存在する場合は、解析している時に手動で編集された可能性のある特定のエキソンを示す (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 (exon - stretching)」アルゴリズムにより結び合わせたエキソンの集合を指す。例えば、FL X X X X X X _ g A A A A A _ g B B B B

B__1__N は「ストレッチ」配列の識別番号であり、XXXXXXXX は Incyte プロジェクト識別番号で、gAAAAA は「エキソストレッチ」アルゴリズムが適用されたヒトゲノム配列の GenBank 識別番号である。さらに、gBBBBB は最も近親の GenBank タンパク質相同体の GenBank 識別番号か、または NCBI RefSeq 識別番号であり、N は特定のエキソンを指す（実施例 5 を参照）。RefSeq 配列が「エキソストレッチ」アルゴリズムのタンパク質相同体として使われた場合は、RefSeq 識別子（「NM」、「NP」または「NT」と表示された）は GenBank 識別子（すなわち、gBBBBB）の代わりに使われる場合もある。

【0120】

あるいは、接頭コードは手で編集されたか、ゲノム DNA 配列から予測されたか、または配列解析方法の組み合わせから由来している構成配列を識別する。次の表は構成配列の接頭コードと、接頭コードと関連する同じ配列の分析方法の例を列記する（実施例 4 と 5 を参照）。

接頭コード	解析のタイプまたはプログラムの例
GN、GFG、ENST	GENSCAN (Stanford University CA, USA) または FGENES (Computer Graphics Group, The Sanger Centre, Cambridge, UK) を用いたゲノム配列からのエキソ予測
GH	手動編集されたゲノム配列の解析
EL	ストレッチまたはストレッチゲノム配列（実施例 5 参照）
INCY	ゲノムの EST 配列のマッピングからの完全長写とエキソ予測ゲノムの位置と EST 構造データを組み合わせでエキソおよびその結果生じる完全長写を予測する

【0121】

場合によっては、最終コンセンサスポリヌクレオチド配列を確認するための列 5 に示すような配列の適用範囲と重複する Incyte cDNA の適用範囲が得られたが、関連する Incyte cDNA 識別番号は示さなかった。

【0122】

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

10

20

30

40

50

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。

【0123】

本発明には、GCREC の変異体も含まれる。好適な GCREC の変異体のアミノ酸配列は、GCREC 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、GCREC 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。

【0124】

本発明には、GCREC をコードするポリヌクレオチドも含まれる。或る例では、GCREC をコードする SEQ ID NO: 20 - 38 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。SEQ ID NO: 20 - 38 のポリヌクレオチド配列は、配列表に示されているように等価 RNA 配列をも含むが、窒素塩基チミンの出現はウラシルに置換され、糖のバックボーンはデオキシリボースではなくリボースから構成されている。

【0125】

本発明には、GCREC 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。具体的には、そのようなポリヌクレオチド配列の変異配列は、GCREC をコードするポリヌクレオチド配列と少なくとも約 70%、或いは少なくとも約 85%、または少なくとも約 95% もの一致率を有する。本発明の特定の実施形態は、SEQ ID NO: 20 - 38 からなる一群から選択された核酸配列と少なくとも 70% のポリヌクレオチド配列同一性、或いは少なくとも 85% のポリヌクレオチド配列同一性、更には少なくとも 95% ものポリヌクレオチド配列同一性を有する SEQ ID NO: 20 - 38 からなる群から選択された配列を含むポリヌクレオチド配列の変異配列を提供する。上記したポリヌクレオチド変異配列は何れも、GCREC の機能的或いは構造的特徴の少なくとも 1 つを有するアミノ酸配列をコードする。

【0126】

当業者であれば、遺伝暗号の縮重の結果、GCREC 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 GCREC 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。

【0127】

GCREC 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ентな条件下で天然の GCREC のヌクレオチド配列とハイブリダイズ可能であるが、GCREC 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なく GCREC 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 RNA 転写物の作製がある。

【0128】

本発明には、GCREC、GCREC 誘導体及びこれらの断片をコードする DNA 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 GCREC 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。

【0129】

更に本発明には、種々のストリンジェントな条件下で、請求項に記載されたポリヌクレオチド配列、特に、SEQ ID NO: 20 - 38及びそれらの断片とハイブリダイズ可能なポリヌクレオチド配列が含まれる(例えば、Wahl, G. M. 及びS. L. Berger (1987) *Methods Enzymol.* 152: 399 - 407、Kimmel, A. R. (1987) *Methods Enzymol.* 152: 507 - 511を参照)。アニーリング及び洗浄条件を含むハイブリダイゼーションの条件は、「定義」に記載されている。

【0130】

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)。

【0131】

GREC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1つである制限部位PCR法は、ユニバーサルプライマー及びネステッドプライマーを用いてクローニングベクター内のゲノムDNAから未知の配列を増幅する方法である (Sarkar, G. (1993) *PCR Methods Applic* 2: 318 - 322等を参照)。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鑄型から未知の配列を増幅する方法である。鑄型は、既知のゲノム遺伝子座及びその周辺の配列からなる制限断片から得る (Triglia, T.ら (1988) *Nucleic Acids Res* 16: 8186等を参照)。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.ら (1991) *PCR Methods Applic* 1: 111 - 119等を参照)。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: 3055 - 3060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/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。

【0132】

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

【0133】

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。キャピラリー電気泳動法は、特定のサンプルに少量しか存在しないような DNA 小断片のシーケンシングに特に適している。

【0134】

本発明の別の実施例では、GC REC 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 GC REC、GC REC 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 GC REC の発現に利用し得る。

【0135】

種々の目的で GC REC 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。

【0136】

本発明のヌクレオチドは、Molecular BreedingTM (Maxygen Inc., Santa Clara CA. 米国特許第 5,837,458 号、Chang, C.-C. 氏 (1999) Nat. Biotechnol. 17:793-797、Christians, F.C. 氏 (1999) Nat. Biotechnol. 17:259-264、Cramer, A. 氏 (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、GC REC 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る。次にこれらの好適な変異体をプールし、更に反復してDNAシャッフリング及び選択/スクリーニングを行ってもよい。このように、遺伝の多様性は「人為的」品種改良及び急速な分子の進化を経て創生される。例えば、ランダムポイント突然変異を有する単一の遺伝子の断片を組み換えて、スクリーニングし、その後所望の特性が最適化されるまでシャッフリングすることができる。或いは、所定の遺伝子を同種または異種のいずれかから得た同一遺伝子ファミリーの相同遺伝子と組み換えて、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。

【0137】

別の実施例によれば、当分野でよく知られている化学的方法を用いて、GCRECをコードする配列の全部或いは一部を合成し得る(Caruthers, M. H.ら(1980) *Nucleic Acids Symp. Ser. 7: 215-223*, Horn, T.ら(1980) *Nucleic Acids Symp. Ser. 7: 225-232*等を参照)。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(Perkin Elmer)を用いて達成し得る。更にGCRECのアミノ酸配列またはその任意の一部は、直接合成する間、または他のタンパク質から得た配列またはその任意の一部と結合する間、或いはその両方を行っている間に変更し、変異型ポリペプチドを生成することが可能である。

【0138】

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

【0139】

生物学的に活性なGCRECを発現させるために、GCRECまたはその誘導体をコードするヌクレオチド配列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。これらの必要な要素には、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含まれる。このような要素は、長さ及び特異性が様々である。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も可能である。このようなシグナルには、ATG開始コドンと、コザック配列などの近傍の配列が含まれる。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。(例えば、Scharf, D. 他(1994) *Results Probl. Cell Differ.* 201-18-162.を参照)。

【0140】

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 Cloning. A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, 4章及び8章, 及び16-17章; 及び Ausubel, F.M. 他. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9章及び13章1-4章を参照)。

【0141】

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. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509、; Engelhard, E.K. ら (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227、Sandig, V. ら (1996) *Hum. 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頁、Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659、Harrington, J.J. ら (1997) *Nat. Genet.* 15:345-355等を参照)レトロウイルス、アデノウイルス、ヘルペスウイルスまたはワクシニアウイルス由来の発現ベクター、または種々の細菌性プラスミド由来の発現ベクターを用いて、ヌクレオチド配列を標的器官、組織または細胞集団へ輸送することができる(Di Nicola, M. ら (1998) *Cancer Gen. Ther.* 5(6):350-356、Yu, M. ら (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344、Buller, R.M. ら (1985) *Nature* 317(6040):813-815; McGregor, D.P. ら (1994) *Mol. Immunol.* 31(3):219-226、Verma, I.M. and N. Somia (1997) *Nature* 389:239-242等を参照)。本発明は使用される宿主細胞によって限定されるものではない。

【0142】

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) 40

J. Biol. Chem. 264:55035509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T5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。

【0143】

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、酵母菌サッカロミセス・セレビジエまたは Pichia pastoris に使用可能である。更に、このようなベクターは、発現したタンパク質の分泌か細胞内への保持のどちらかを誘導し、安定した増殖のために宿主ゲノムの中に外来配列を組み込む。(例えば、Ausubel, 1995, 前出、Bitter, G. A. ら (1987) *Methods Enzymol.* 153:516-544、及びScorer, C. A. ら (1994) *Bio/Technology* 121-181-184. を参照)。

【0144】

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、CaMV由来の35S及び19Sプロモーターを単独で、或いは、TMV(タカマツ, 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て促進し得る。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形質転換によって、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頁等を参照)。

【0145】

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/翻訳複合体に連結され得る。可欠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. および T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:36553659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。

【0146】

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。(例えば、Harrington, J. J. 他 (1997) *Nat Genet.* 15:345-355を参照)。

【0147】

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。

【0148】

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*⁻単純細胞のために用いられるヘルペスウイルス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; 及び Lowy, I. 他(1980) Cell 22: 817-823を参照)。また、選択の基礎として代謝拮抗物質、抗生物質或いは除草剤への耐性を用いることができる。例えば *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. ら(1980) Proc. Natl. Acad. Sci. USA 77: 35673570; Colbere Garapin, F. ら(1981) J. Mol. Biol. 150: 114等を参照。)この他の選択可能な遺伝子、例えば、代謝のための細胞の必要条件を変える *trpB* 及び *hisD* は、文献に記載されている(Hartman, S. C. and 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: 121131等を参照)。

10

20

【0149】

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をコードする配列をマーカー遺伝子配列内に挿入すると、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。誘導または選択に应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

【0150】

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。

40

【0151】

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. ら(1990) Serological Methods, a Laboratory Ma

50

n u a l . A P S P r e s s . S t P a u l . M N , S e c t . I V , C o l i g a n , J . E . 5 (1 9 9 7) C u r r e n t P r o t o c o l s i n I m m u n o l o g y , G r e e n e P u b . A s s o c i a t e s a n d W i l e y - I n t e r s c i e n c e , N e w Y o r k N Y , P o u n d , J . D . (1 9 9 8) I m m u n o c h e m i c a l P r o t o c o l s , H u m a n s P r e s s , T o t o w a N J等を参照)。

【0152】

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、または標識ヌクレオチドを用いるPCR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

【0153】

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の直接分泌を誘導するシグナル配列を含むように設計し得る。

【0154】

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

【0155】

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はフェニルアルシノオ

40

50

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、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の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、GCRECが精製後に異種部分から切断できるように、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切断部位を含めるように融合タンパク質を遺伝子操作することもできる。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。

【0156】

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。

10

【0157】

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。

20

【0158】

一実施例では、このように同定された化合物は、例えばリガンドやその断片などのGCRECの天然のリガンド、または天然の基質、構造的または機能的な擬態性または自然結合パートナーに密接に関連している(Coligan, J. E. 他(1991) Current Protocols in Immunology 1(2)の5章等を参照)。同様に、化合物は、GCRECが結合する天然受容体、或いは例えばリガンド結合部位などの少なくとも受容体のある断片に密接に関連し得る。何れの場合も、既知の技術を用いてこの化合物を合理的に設計することができる。一実施例では、このような化合物に対するスクリーニングには、分泌タンパク質或いは細胞膜上のタンパク質の何れか一方としてGCRECを発現する好適な細胞の作製が含まれる。好適な細胞には、哺乳動物、酵母、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。

30

【0159】

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、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。

40

【0160】

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

50

件下でGCRECを含む*in vitro*または細胞遊離系と結合させてアッセイを実施する。これらアッセイの何れかにおいて、GCRECの活性を調整する試験化合物は間接的に結合することが可能であり、試験化合物と直接接触する必要がない。少なくとも1つから複数の試験化合物をスクリーニングすることができる。

【0161】

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

【0162】

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）。

【0163】

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

【0164】

（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タンパク質共役受容体との間に存在する。さらに、GCRECの発現は脳組織、胎性脳組織、大腸ポリープ、罹患大腸組織、大腸腫瘍組織、罹患胆嚢組織、心臓組織、罹患乳房組織、インターロイキン-5刺激性好酸球、腫瘍組織、および生殖組織に密接に関連している。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/炎症性疾患及び代謝障害並びにウイルス感染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

50

【0165】

従って、或る実施例において、G C R E Cの発現または活性の低下に関連した疾患の治療または予防のために、患者にG C R E Cまたはその断片や誘導体を投与することが可能である。限定するものではないが、そのような疾患のうち、細胞増殖異常には日光性角化症及びアテローム性動脈硬化、滑液包炎、硬変、肝炎、混合型結合組織病(M C T D)、骨髓線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症、並びに腺癌及び白血病、リンパ腫、黒色腫、骨髄腫、肉腫、及び奇形癌、具体的には、副腎、膀胱、骨、骨髄、脳、乳房、頸部、胆嚢、神経節、消化管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、子宮の癌が含まれ、

10

【0166】

神経の疾患の中には、癲癇、虚血性脳血管障害、脳卒中、大脳新生物、アルツハイマー病、ピック病、ハンチントン病、痴呆、パーキンソン病及びその他の錐体外路障害、筋萎縮性側索硬化及びその他の運動ニューロン障害、進行性神経性筋萎縮症、色素性網膜炎、遺伝性運動失調、多発性硬化症及び他の脱髄疾患、細菌性及びウイルス性髄膜炎、脳膿瘍、硬膜下蓄膿症、硬膜外膿瘍、化膿性頭蓋内血栓性静脈炎、脊髄炎及び神経根炎、ウイルス性中枢神経系疾患と、クールー及びクロイツフェルト ヤコブ病、G e r s t m a n n - S t r a u s s l e r - S c h e i n k e r症候群を含むプリオン病と、致死性家族性不眠症、神経系性栄養病及び代謝病、神経線維腫症、結節硬化症、小脳網膜血管芽腫(c e r e b e l l o r e t i n a l h e m a n g i o b l a s t o m a t o s i s)、脳3叉

20

【0167】

心血管疾患の中には、動静脈瘤、アテローム硬化、高血圧、脈管炎、レイノー病、静脈奇形、動脈解離、静脈瘤、血栓静脈炎及び静脈血栓、血管の腫瘍、血栓崩壊の合併症、バルーン血管形成術、血管置換術、バイパス手術、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、僧帽弁逸脱、リウマチ熱、リウマチ性心疾患、感染性心内膜炎、非細菌性血栓性心内膜炎、全身性エリテマトーデスの心内膜炎、カルチノイド心疾患、心筋症、心筋炎、心膜炎、腫瘍性心疾患、先天性心臓疾患、及び心臓移植の合併症が含まれ、

30

【0168】

また、胃腸疾患が含まれ、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、腹部アンギナ、胸焼け、胃腸炎、イレウス、腸管感染、消化性潰瘍、胆石症、胆嚢炎、胆汁うっ滞、膵臓炎、膵臓癌、胆道疾患、肝炎、高ビリルビン血症、硬変症、肝臓の受動性うっ血、ヘパトーム、感染性大腸炎、潰瘍性大腸炎、潰瘍性直腸炎、クローン病、ホイップル病、マロリー ヴァイス症候群、結腸癌、結腸閉塞、過敏性腸症候群、短小腸症候群、下痢、便秘、胃腸出血、後天性免疫不全症候群(A I D S)腸症、黄疸、肝性脳症、肝腎症候群、肝炎、血色素症、ウィルソン病、₁ アンチトリプシン欠損症、ライ症候群、原発性硬化性胆管炎、肝梗塞、門脈循環閉塞及び血栓、小葉中心壊死、肝臓紫斑病、肝静脈血栓、肝静脈閉塞症、子癇前症、子癇、妊娠性急性肝脂肪、妊娠性肝臓内胆汁うっ滞と、結節性再生及び腺腫、癌腫を含む肝癌が含まれ、

40

【0169】

50

自己免疫/炎症性の疾患の中には、後天性免疫不全症候群（AIDS）及びアジソン病、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、自己免疫性多腺性内分泌カンジダ性外胚葉ジストロフィー（APCED）、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、糖尿病、肺気腫、リンパ球毒素性一時性リンパ球減少症、赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性大腸症候群、多発性硬化症、重症筋無力症、心筋または心膜炎、骨関節炎、骨粗しょう症、脾炎、多発性筋炎、乾癬、ライター症候群、リウマチ様関節炎、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、血小板減少性紫斑病、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌合併症、血液透析、体外循環、ウイルス感染症、細菌感染症、真菌感染症、寄生虫感染症、原虫感染症、蠕虫感染症、外傷が含まれ、

10

【0170】

また、糖尿病、肥満症および骨粗しょう症などの代謝異常、およびアデノウイルス、アレナウイルス、ブニavirus、カルチウイルス、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、パポウイルス、パラミキソウイルス、ピコルナウイルス、ポックスウイルス、レオウイルス、レトロウイルス、ラブドウイルス、およびトガウイルスと分類されるウイルス性媒介物による感染が含まれる。

【0171】

別の実施例では、GCREC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GCREC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

【0172】

更に別の実施例では、実質的に精製されたGCREC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GCREC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。

【0173】

更に別の実施例では、GCREC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GCREC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。

30

【0174】

更に別の実施例では、患者にGCRECのアンタゴニストを投与して、GCREC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/炎症疾患及び代謝異常並びにウイルス感染がある。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GCREC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GCRECと特異結合する抗体を用いることができる。

【0175】

別の実施例では、GCREC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GCREC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。

40

【0176】

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す得る。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。

50

【0177】

G C R E C のアntagオニストは、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。

【0178】

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、G C R E C、または G C R E C の任意の断片またはオリゴペプチドであって免疫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、スカシガイのヘモシニアン、ジニトロフェノール等の界面活性剤とがある。ヒトに用いられるアジュバントの中では、B C G（カルメット ゲラン桿菌）及びコリネバクテリウム パルヴムが特に好ましい。

【0179】

G C R E C 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り且つ小さな天然の分子の全アミノ酸配列を含むことが望ましい。G C R E C アミノ酸の短い伸長部は、別のタンパク質、例えばスカシガイの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。

【0180】

G C R E C 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）。

【0181】

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、G C R E C 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）。

【0182】

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)。

【0183】

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 256: 1275 - 1281 等を参照)。

【0184】

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参与している。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もよい(前出のPoundの文献)。

【0185】

ラジオイムノアッセイ技術と共に様々な方法、例えばスキャ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Kaで表す。Ka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Ka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Kaは、親和性の真の測定値を表す。Ka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。Ka値が約 $10^6 \sim 10^7$ L/molの範囲にあるような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及び類似の処理に用いるのが好ましい(Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington, DC、Liddell, J. E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY)。

【0186】

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1~2 mg/ml、好ましくは5~10 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)。

【0187】

10

20

30

40

50

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うことができる。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、修飾ヌクレオチド)を設計して遺伝子発現を変更することができる。このような技術は当分野では周知であり、センスまたはアンチセンスオリゴヌクレオチドまたは大きな断片が、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., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJを参照。)

【0188】

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及びScanlon, K.J.ら(1995)9(13):1288-1296.等を参照)アンチセンス配列はまた、例えばレトロウイルスやアデノ関連ウイルスベクター等のウイルスベクターを用いて細胞内に導入することもできる(Miller, A.D. (1990) *Blood* 76:271、前出のAusubel, Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347等を参照)。その他の遺伝子送達機構には、リポソーム系、人工的なウイルスエンベロ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)。

【0189】

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染色体関連遺伝(Cavazzana-Calvo, M.他(2000) *Science* 288:669-672)により特徴付けられる重度の複合型免疫欠損症(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症(Blaese, R.M.他(1995) *Science* 270:475-480、Bordignon, C.他(1995) *Science* 270:470-475)、嚢胞性繊維症(Zabner, J.他(1993) *Cell* 75:207-216; Crystal, R.G.他(1995) *Hum. Gene Therapy* 6:643-666、Crystal, R.G.他(1995) *Hum. Gene Therapy* 6:667-703)、サラセミア(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. および 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)細胞内の寄生生物(例えばヒト免疫不全ウイルス(HIV)(Baltimore, D. (1988) *Nature* 335:395-396、Poeschl, E.他(1996) *Proc. Natl. Acad. Sci. USA.* 93:11395-11399)、B型若しくはC型肝炎ウイルス(HBV、HCV)、Candida albicans及びParacoccidioides brasiliensis等の真菌寄生菌、並びに熱帯熱マラリア原虫及びクルーズトリパノソーム等の原虫寄生生物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

10

20

30

40

50

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。

【0190】

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 vitro* の細胞に用いる機械的導入技術には、(i) 個々の細胞内への直接的なDNA微量注射法、(ii) 遺伝子銃、(iii) リポソームを介した形質移入、(iv) 受容体を介した遺伝子導入、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。

10

【0191】

GCRECの発現に効果的である発現ベクターには、限定するものではないが、PCDNA3.1、EPITAG、PRCCMV2、PREP、PVAX、PCR2-TOPOTAベクター(Invitrogen, Carlsbad CA)、PCMVS-CRIPT、PCMV-TAG、PEGSH/PERV(Stratagene, La Jolla CA)およびPTETOFF、PTETON、PTRE2、PTRE2-LUC、PTKHYG(Clontech, Palo Alto CA)が含まれる。GCRECは、(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)またはアクチン遺伝子)、(ii) 誘導性プロモーター(例えばテトラサイクリン調節性プロモーター(Gossen, M. and H. Bujard (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89: 5547-5551、Gossen, M. ら (1995) *Science* 268: 1766-1769、Rossi, F.M.V. and H.M. Blau (1998) *Curr. Opin. Biotechnol.* 9: 451-456)、市販のInvitrogen社のT-REXプラスミドに含まれる)、エクジソン誘導性プロモーター(Invitrogen社のプラスミドPVGRXR及びPINDから得られる)、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(前出のRossi, F.M.V. and H.M. Blau)、または(iii) 正常個体由来の、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て、発現させることができる。

20

30

【0192】

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. および 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。

40

【0193】

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

50

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（VPC L）において増殖され、VPC Lは、標的細胞上の受容体に対する親和性を有するエンベロープ遺伝子またはVSVg等の汎親和性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. 61:1639-1646、Adam, M.A. and 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）。

【0194】

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脾臓の無損傷の脾島内に導入するために可変性であることが証明された（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 vector 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 及び Verma, I.M. および N. Somia (1997) Nature 18:389:239-242も参照されたい。両文献は、引用することをもって本明細書の一部とする。

【0195】

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が親和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 swa

ins for gene transfer”)に開示されており、該特許の引用をもって本明細書の一部とする。米国特許第5,804,413号には、ヒト遺伝子治療を含む目的のために好適なプロモーターの制御下において細胞に導入される少なくとも1つの外在性遺伝子を有するゲノムを含む組換えHSV d92についての記載がある。上記特許はまた、ICP4、ICP27及びICP22のために除去される組換えHSV系統の作製及び使用について開示している。HSVベクターについては、Goins, W.F.ら(1999) *J. Virol.* 73:519-532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。

10

【0196】

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キャプシッ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。

20

30

【0197】

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頁等を参照*)。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。

40

【0198】

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、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とその後に起こる内ヌクレオチド鎖切断に関

50

与している。例えば、遺伝子操作で作られたハンマーヘッド型リボザイム分子は、G C R E Cをコードする配列の内ヌクレオチド鎖分解性の切断を特異的且つ効果的に触媒する可能性がある。

【0199】

任意のRNA標的内の特異的リボザイム切断部位を、GUA、GUU、GUC配列を含めたりリボザイム切断部位に対する標的分子をスキャンすることによって先ず同定する。一度同定すると、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配列が、そのオリゴヌクレオチドを機能不全にするような2次構造の特徴をもっていないかを評価することが可能になる。候補標的の適合性の評価も、リボヌクレアーゼ保護アッセイを用いて相補的オリゴヌクレオチドとのハイブリダイゼーションの実施容易性をテストすることによって行うことができる。

10

【0200】

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学合成のように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、G C R E C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、細胞系、細胞または組織内に導入することができる。

20

【0201】

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、クエオシン(queosine)、ワイプトシン(wybutosine)等を加えることでできる。

30

【0202】

本発明の追加実施例は、G C R E C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発現変異を起こすのに有効な化合物には、オリゴヌクレオチド、アンチセンスオリゴヌクレオチド、三重らせん形成オリゴヌクレオチド、転写因子その他のポリペプチド転写制御因子、及び特異ポリヌクレオチド配列と相互作用し得る非高分子化学的実体がある。有効な化合物は、ポリヌクレオチド発現のインヒビターまたはエンハンサーのいずれかとして作用することによりポリヌクレオチド発現を変異し得る。従って、G C R E C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、G C R E C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、G C R E C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、G C R E Cをコードするポリヌクレオチドの発現を特異的に促進する化合物が治療上有用であり得る。

40

【0203】

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。G C R E Cをコー

50

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。G C R E C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、G C R E C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)。

【0204】

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

【0205】

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。

【0206】

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、G C R E C、G C R E C に対する抗体、擬態、アゴニスト、アンタゴニスト、または G C R E C インヒビターから構成し得る。

【0207】

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。

【0208】

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

10

20

30

40

50

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。

【0209】

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。

【0210】

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ポソーム製剤は、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 タンパク質から陽イオン 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)。

10

【0211】

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

20

【0212】

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) 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。

30

【0213】

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、患者の通常の状態、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、特定の薬剤の半減期及びクリアランス率によって 3 ~ 4 日毎に 1 度、1 週間に 1 度、或いは 2 週間に 1 度の間隔で投与し得る。

40

【0214】

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 1 g 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。

【0215】

(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

50

G C R E C や G C R E C 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、G C R E C 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。G C R E C 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 G C R E C 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。

【0216】

G C R E C を測定するための様々なプロトコル、例えば E L I S A、R I A、F A C S 等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、制御、検体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。

10

【0217】

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 R N A 及び D N A 分子、そして P N A 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。

20

【0218】

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 P C R 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ェ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。

30

【0219】

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、D N A あるいは R N A が可能であり、S E Q I D N O : 20 - 38 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。

【0220】

G C R E C をコードする D N A 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、m R N A 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。m R N A プローブ作製のためのベクターは、当業者に知られており、市販されており、好適な R N A ポリメラーゼ及び好適な標識されたヌクレオチドを加えることによって、i n v i t r o で R N A プローブを合成するために用いられ得る。ハイブリダイゼーションプローブは、種々のレポーターの集団によって標識され得る。レポーター集団の例としては、³²P または ³⁵S 等の放射性核種、或いはアビジン/ビオチン結合系を介してプローブに結合されたアルカリホスファターゼ等の酵素標識などが挙げられる。

40

【0221】

50

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、そのような疾患のうち、細胞増殖異常の中には日光性角化症及びアテローム性動脈硬化、滑液包炎、硬変、肝炎、混合型結合組織病 (M C T D)、骨髄線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症、並びに腺癌及び白血病、リンパ腫、黒色腫、骨髄腫、肉腫、及び奇形癌、具体的には、副腎、膀胱、骨、骨髄、脳、乳房、頸部、胆嚢、神経節、消化管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、子宮の癌が含まれ、

【0222】

神経の疾患の中には、癲癇、虚血性脳血管障害、脳卒中、大脳新生物、アルツハイマー病、ピック病、ハンチントン病、痴呆、パーキンソン病及びその他の錐体外路障害、筋萎縮性側索硬化及びその他の運動ニューロン障害、進行性神経性筋萎縮症、色素性網膜炎、遺伝性運動失調、多発性硬化症及び他の脱髄疾患、細菌性及びウイルス性髄膜炎、脳膿瘍、硬膜下蓄膿症、硬膜外膿瘍、化膿性頭蓋内血栓性静脈炎、脊髄炎及び神経根炎、ウイルス性中枢神経系疾患と、クールー及びクロイツフェルト ヤコブ病、Gerstmann-Straussler-Scheinker 症候群を含むプリオン病と、致死性家族性不眠症、神経系性栄養病及び代謝病、神経線維腫症、結節硬化症、小脳網膜血管芽腫 (cerebelloretinal hemangioblastomatosis)、脳3叉神経血管症候群、ダウン症を含む中枢神経系性精神遅滞及び他の発達障害、脳性麻痺、神経骨格異常症、自律神経系障害、脳神経疾患、脊髄疾患筋ジストロフィー、および他の神経筋障害、末梢神経系疾患、皮膚筋炎及び多発性筋炎と、遺伝性、代謝性、内分泌性、及び中毒性ミオパシーと、重症筋無力症、周期性四肢麻痺と、気分性及び不安性の障害、分裂病性疾患を含む精神病と、季節性障害 (SAD) と、静座不能、健忘症、緊張病、糖尿病性ニューロパシー、錐体外路性終末欠陥症候群 (遅発性ジスキネジア)、ジストニー、分裂病性精神障害、帯状疱疹後神経痛、及びトゥレット病と、進行性核上麻痺、大脳皮質基底核変性症及び家族性前頭側頭性健忘症が含まれ、

【0223】

心血管疾患の中には、動静脈瘤、アテローム硬化、高血圧、脈管炎、レイノー病、静脈奇形、動脈解離、静脈瘤、血栓静脈炎及び静脈血栓、血管の腫瘍、血栓崩壊の合併症、バルーン血管形成術、血管置換術、バイパス手術、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、僧帽弁逸脱、リウマチ熱、リウマチ性心疾患、感染性心内膜炎、非細菌性血栓性心内膜炎、全身性エリテマトーデスの心内膜炎、カルチノイド心疾患、心筋症、心筋炎、心膜炎、腫瘍性心疾患、先天性心臓疾患、及び心臓移植の合併症が含まれ、

【0224】

また、胃腸疾患が含まれ、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、腹部アンギナ、胸焼け、胃腸炎、イレウス、腸管感染、消化性潰瘍、胆石症、胆嚢炎、胆汁うっ滞、膵臓炎、膵臓癌、胆道疾患、肝炎、高ビリルビン血症、硬変症、肝臓の受動性うっ血、ヘパトーム、感染性大腸炎、潰瘍性大腸炎、潰瘍性直腸炎、クローン病、ホイップル病、マロリー ヴァイス症候群、結腸癌、結腸閉塞、過敏性腸症候群、短小腸症候群、下痢、便秘、胃腸出血、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、 γ_1 アンチトリプシン欠損症、ライ症候群、原発性硬化性胆管炎、肝梗塞、門脈循環閉塞及び血栓、小葉中心壊死、肝臓紫斑病、肝静脈血栓、肝静脈閉塞症、子癇前症、子癇、妊娠性急性肝脂肪、妊娠性肝臓内胆汁うっ滞と、結節性再生及び腺腫、癌腫を含む肝癌が含まれ、

【0225】

自己免疫 / 炎症性の疾患の中には、後天性免疫不全症候群 (AIDS) 及びアジソン病、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-

ム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫性多腺性内分泌カンジダ性外胚葉ジストロフィー（A P E C E D）、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、糖尿病、肺気腫、リンパ球毒素性一時性リンパ球減少症、赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性大腸症候群、多発性硬化症、重症筋無力症、心筋または心膜炎、骨関節炎、骨粗しょう症、膵炎、多発性筋炎、乾癬、ライター症候群、リウマチ様関節炎、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、血小板減少性紫斑病、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌合併症、血液透析、体外循環、ウイルス感染症、細菌感染症、真菌感染症、寄生虫感染症、原虫感染症、蠕虫感染症、外傷が含まれ、

10

【0226】

また、糖尿病、肥満症および骨粗しょう症などの代謝異常、およびアデノウイルス、アレナウイルス、ブニavirus、カルチウイルス、コロナウイルス、フィロウイルス、ヘパドナウイルス、ヘルペスウイルス、フラビウイルス、オルソミクソウイルス、パルボウイルス、パポバウイルス、パラミキソウイルス、ピコルナウイルス、ポックスウイルス、レオウイルス、レトロウイルス、ラウドウイルス、およびトガウイルスと分類されるウイルス性媒介物による感染が含まれる。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（d i p s t i c k）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。

20

【0227】

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制御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。

30

【0228】

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。

40

【0229】

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。

【0230】

癌に関しては、個体からの生体組織における異常な量の転写物（過少発現または過剰発現）の存在は、疾患の発生素質を示したり、実際に臨床的症状が現れる前に疾患を検出する方法を提供したりし得る。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

50

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。

【0231】

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 の利用に關与し得る。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。

10

【0232】

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チドプライマーを用いて単1塩基多型性 (S N P) を検出し得る。S N P は、多くの場合にヒトの先天性または後天性遺伝病の原因となるような置換、挿入及び欠失である。限定するものではないが S N P の検出方法には、S S C P (*single-stranded conformation polymorphism*) 及び蛍光 S S C P (*f S S C P*) 法がある。S S C P では、G C R E C をコードするポリヌクレオチド配列由来のオリゴヌクレオチドプライマーを用いて、ポリメラーゼ連鎖反応法 (P C R) を用いた D N A の増幅を行う。D N A は例えば、病変組織または正常組織、生検サンプル、体液その他に由来し得る。D N A 内の S N P は、一本鎖形状の P C R 生成物の2次及び3次構造に差異を生じさせる。差異は非変性ゲル中でのゲル電気泳動法を用いて検出可能である。f S C P では、オリゴヌクレオチドプライマーを蛍光性に標識する。それによって D N A シークエンシング機などの高処理機器でアンプリマー (*amplimer*) の検出が可能になる。更に、インシリコ S N P (*in silico SNP, is SNP*) と呼ばれる配列データベース分析法は、一般的なコンセンサス配列に配列されるような個々の重畳する D N A 断片の配列を比較することにより、多型性を同定し得る。これらのコンピュータベースの方法は、D N A の実験室での調整及び統計モデル及び D N A 配列クロマトグラムの自動分析を用いたシークエンシングのエラーに起因する配列の変異をフィルタリングして除去する。別の態様では、例えば高処理 M A S S A R R A Y システム (*Sequenom, Inc., San Diego CA*) を用いた質量分析により S N P を検出し、特徴付ける。

20

30

【0233】

G C R E C 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. 他 (1993) *J. Immunol. Methods* 159: 235244; Duplaa, C. 他 (1993) *Anal. Biochem.* 212: 229236 を参照。) 目的のオリゴマーが種々の希釈液中に存在し、分光光度法または比色反応によって定量が迅速になるような高処理フォーマットのアッセイを行うことによって、複数のサンプルの定量速度を加速することができる。

40

【0234】

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、突然変異及び多型性の同定に用いることができる。この情報を用いることで、遺伝子機能を決定し、疾患の遺伝的根拠を理解し、疾患を診断し、遺伝子発現の機能としての疾病の進行/後退をモニターし、疾病治療における薬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

50

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ィー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ィールに基づき、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。

【0235】

別の実施例では、G C R E C に特異的な抗体、G C R E C 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ィールをモニターまたは測定することが可能である。

【0236】

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、所定の条件下で所定の時間に発現した遺伝子の数及び相対存在量を定量することにより分析される (Seilliamer 他、米国特許第 5, 840, 484 号の "Comparative Gene Transcript Analysis" を参照。この特許に言及することを以って本明細書の一部とする)。従って、特定の組織または細胞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。

【0237】

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

【0238】

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、in vitro モデル系及び薬剤の前臨床評価に関連して使用し得る。全ての化合物は、作用及び毒性のメカニズムを暗示する、しばしば分子フィンガープリントまたは毒性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. および 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同一のシグネチャを有する場合には、毒性特性を共有している可能性がある。フィンガープリントまたはシグネチャは、多数の遺伝子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は、最も有用且つ正確である。理想的には、発現をゲノム全域にわたって測定すると、最高品質のシグネチャが得られる。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も同様に重要であり、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を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、毒性シグネチャを用いた中毒学的スクリーニングの際に、全ての発現した遺伝子配列を含めることは重要且つ望ましいことである。

【0239】

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。

【0240】

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。

10

20

【0241】

タンパク質の(proteomic)プロフィールは、GCRCに特異的な抗体を用いてGCRC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. 5. (1999) Anal. Biochem. 270: 103-111, Mendozze, L. G. 5. (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。

30

【0242】

プロテオームレベルでの毒性サインも中毒学的スクリーニングに有用であり、転写レベルでの毒性サインと並行に分析するべきである。或る組織の或るタンパク質に対しては、転写とタンパク質の存在量の相関が乏しいこともあるので(Anderson, N. L. 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えるような化合物の分析においてプロテオーム毒性サインは有用たり得る。更に、体液中での転写の分析はmRNA急速な分解により困難であるので、タンパク質のプロフ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。

40

【0243】

別の実施例では、タンパク質を含有する生物学的サンプルを試験化合物で処理することにより試験化合物の毒性を算定する。処理された生物学的サンプル中で発現したタンパク質は、各タンパク質の量を定量し得るように分離する。各タンパク質の量を、非処理生物学

50

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。

【0244】

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

10

【0245】

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

20

【0246】

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用なハイブリダイゼーションプローブを作製するために、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でのコード配列の保存により、染色体マッピング中に望ましくないクロスハイブリダイゼーションが生じる可能性がある。核酸配列は、特定の染色体、染色体の特定領域または人工形成の染色体、例えば、ヒト人工染色体 (HAC)、酵母人工染色体 (YAC)、細菌人工染色体 (BAC)、細菌P1産物、或いは単一染色体cDNAライブラリに対してマッピングされる。(例えば、Harrington, J. J. 他 (1997) Nat. Genet. 15:345-355; Price, C. M. (1993) Blood Rev. 7:127134; and Trask, B. J. (1991) Trends Genet. 7:149154を参照。) 一度マッピングすると、本発明の核酸配列を用いて例えば病状の遺伝を特定の染色体領域の遺伝または制限酵素断片長多型 (RFLP) と関連させるような遺伝子連鎖地図を作製できる。(例えば、Lander, E. S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357を参照。)

30

40

蛍光原位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。

【0247】

50

確定した染色体マーカーを用いた結合分析等の物理的マッピング技術及び染色体標本原位ハイブリッド形成法を用いて、遺伝地図を拡張することができる。例えばマウスなど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。疾患または症候群が、血管拡張性失調症の11q22-23領域等、特定の遺伝子領域への遺伝的結合によって大まかに位置決めがなされると、該領域に対するいかなるマッピングも、更なる調査のための関連遺伝子或いは調節遺伝子を表すことができる(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。

10

【0248】

本発明の別の実施例では、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、GCREC、GCREC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。GCRECとテストされる薬剤との結合複合の形成は計測できる。

【0249】

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(Geyssen,らの(1984)PCT出願番号WO84/03564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、GCREC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GCRECを検出する。精製したGCREC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。

20

【0250】

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、GCRECを結合できる中和抗体が、GCREC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若しくは数個の抗原決定因子をGCRECと共有するペプチドの存在を検出する。

30

【0251】

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、まだ発展途上である分子生物学技術のいずれにおいても、GCRECをコードするヌクレオチド配列を用いることができる。

【0252】

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

40

にも本発明を限定するものではない。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60/221,478号、第60/223,268号、第60/231,121号、第60/232,691号、第60/235,146号、第60/227,054号及び第60/232,243号は、言及することをもって本明細書の一部となす。

【0253】

(実施例)

1 cDNAライブラリの作製 Incyte cDNAは、LIFESEQ GOLDデータベース(Incyte Genomics, Palo Alto CA)に記載されたcDNAライブラリ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

50

グアニジニウムイソチオシアネート溶液に溶解した組織もあり、また、ホモジナイズしてフェノールまたは好適な変性剤の混合液に溶解した組織もある。変性剤の混合液は、例えばフェノールとグアニジニウムイソチオシアネートの単相溶液であるTRIzol (Life Technologies) 等である。結果として得られた溶解物は、塩化セシウム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。

【0254】

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

【0255】

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、好適な制限酵素でcDNAを消化した。殆どのライブラリに対して、cDNAのサイズ(300~1000bp)選択は、SEPHACRYL S1000、SEPHAROSE CL2BまたはSEPHAROSE CL4Bカラムクロマトグラフィー(Amersham Pharmacia Biotech)、或いは調製用アガロースゲル電気泳動法を用いて行った。合成オリゴヌクレオチドアダプターを二本鎖cDNAに連結反応させ、好適な制限酵素または酵素でcDNAを消化した。好適なプラスミドは、例えばPBLUESCRIPTプラスミド(Stratagene)、pSPORT1プラスミド(Life Technologies)またはpINCY(INCYTE Pharmaceuticals, Palo Alto CA)等である。組換えプラスミドは、Stratagene社のXL1-Blue、XL1-BIueMRFまたはSOLR、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。

【0256】

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。

【0257】

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. Bioche

10

20

30

40

50

m. 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PICO GREEN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。

【0258】

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Incyte cDNAを、以下に示すようにシークエンシングした。cDNAのシークエンス反応は、標準的方法或いは高処理装置、例えばABI CATALYST 800 サーマルサイクラー (Applied Biosystems) またはPTC-200 サーマルサイクラー (MJ Research) をHYDRAマイクロディスペンサー (Robbins Scientific) またはMICROLAB 2200 (Hamilton) 液体転移システムと併用して処理した。cDNAのシークエンス反応は、Amersham Pharmacia Biotech社が提供する試薬またはABIシークエンシングキット、例えばABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems) に与えられた試薬を用いて準備した。cDNAのシークエンス反応の電気泳動的分離及び標識したポリヌクレオチドの検出には、MEGABACE 1000 DNAシークエンシングシステム (Molecular Dynamics) 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。

【0259】

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Prosit等のデータベース、PFAM等の隠れマルコフモデル (HMM) ベースのタンパク

質ファミリーデータベースに対する問合せによって完全長ポリペプチド配列を分析した。完全長ポリヌクレオチド配列はまた、MACDNASIS PROソフトウェア（日立ソフトウェアエンジニアリング，South San Francisco CA）及びLASERGENEソフトウェア（DNASTAR）を用いて分析した。ポリヌクレオチド及びポリペプチド配列アラインメントは、アラインメントした配列と配列の一致率も計算するMEGALIGNマルチシーケンサアラインメントプログラム（DNASTAR）に組み込まれているようなCLUSTALアルゴリズムによって特定されるデフォルトパラメータを用いて生成する。

【0260】

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2配列間の相同性が高くなる）。

10

【0261】

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: 20 - 38のポリヌクレオチド配列断片の同定にも利用できる。ハイブリダイゼーション及び増幅技術に有用である約20～約4000ヌクレオチドの断片を表4の列4に示した。

20

【0262】

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タンパク質共役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 及びBurge, C. 及び S. Karlin (1998) Cuff. Opin. Struct. Biol. 8: 346 - 354 参照）。プログラムは予測エキソンを連結し、メチオニンから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タンパク質共役受容体をコードするかを決定するために、コードされたポリペプチドをGタンパク質共役受容体のためのPFAMモジュール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タンパク質共役受容体として注釈を付けてきたことにより、潜在的Gタンパク質共役受容体も同定した。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gbpri及びgbhtgと比較した。必要であれば、genpeptからヒットしたトップのBLASTと比較することによりGenscan予測配列を編集し、余分なまたは取り除かれたエキソンなどのGenscanにより予測された配列のエラーを修正する。BLAST分析はまた、任意のIncyte cDNAまたはGenscan予測配列の公共cDNA適用範囲の発見に用いられるので、転写の証拠を提供する。Incyte cDNA適用範囲が利用できる場合には、この情報を用いてGenscan予測配列を修正または確認した。完全長ポリヌクレオチド配列は、実施例3に記載された構築プロセスを用いて、Incyte cDNA配列及び/または公共のcDNA配列でGenscan予測コード配列を構築することにより得た。或いは、完全長ポリヌクレオチド配列は編集または非編集のGenscan予測コード配列に完全に由来する。

30

40

【0263】

5 cDNA配列データを使ったゲノム配列データの構築

50

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

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

【 0 2 6 4 】

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

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

【 0 2 6 5 】

6 G C R E C をコードするポリヌクレオチドの染色体マッピング

S E Q I D N O : 2 0 - 3 8 を構築するために用いた配列を、B L A S T 及び S m i t h - W a t e r m a n アルゴリズムを用いて、I n c y t e L I F E S E Q データベース及び公共のドメインデータベースの配列と比較した。S E Q I D N O : 2 0 - 3 8 と一致するこれらのデータベースの配列を、P h r a p (表 7) などの構築アルゴリズムを使用して、連続及び重複した配列のクラスタに組み入れた。スタンフォード・ヒトゲノムセンター (S H G C)、ホワイトヘッド・ゲノム研究所 (W I G R)、G e n e t h o n 等の公的な情報源から入手可能な放射線ハイブリッド及び遺伝地図データを用いて、クラスタ化された配列が前もってマッピングされたかを測定した。マッピングされた配列がクラスタに含まれている結果、個々の配列番号を含めてそのクラスタの全配列が地図上の位置に割り当てられた。

【 0 2 6 6 】

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

10

20

30

40

50

は、染色体マーカー間の組換え頻度に基づく計測単位である。平均して、1 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「GeneMap99」 (<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。

【0267】

7 ポリヌクレオチド発現の分析

ノーザン分析は、転写された遺伝情報の存在を検出するために用いられる実験技術であり、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。)

BLASTに適用する類似のコンピュータ技術を用いて、GenBankやLifeSeq (Incyte Pharmaceuticals) 等のヌクレオチドデータベースにおいて同一または関連分子を検索する。ノーザン分析は、多数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。

【0268】

【数1】

(BLASTスコア×配列一致率)

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

【0269】

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。

【0270】

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cDNA配列は、ヒト組織から作製されたcDNAライブラリに由来する。各ヒト組織は、以下の生物/組織カテゴリー—即ち心血管系、結合組織、消化器系、胎芽構造、内分泌系、外分泌腺、女性生殖器、男性生殖器、生殖細胞、血液及び免疫系、肝、筋骨格系、神経系、膵臓、呼吸器系、感覚器、皮膚、顎口腔系、非分類性/混合性または尿路の1つに分類される。各カテゴリーのライブラリ数を数えて、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、

10

20

30

40

50

以下の疾患/病状カテゴリ-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-のライブラリ数を数えて、全カテゴリ-の総ライブラリ数で除する。演算結果の割合は、GC REC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。

【0271】

8 GC RECをコードするポリヌクレオチドの伸長

完全長のポリヌクレオチド配列もまた、完全長分子の適切な断片から設計したオリゴヌクレオチドプライマーを用いて該断片を伸長させて生成した。一方のプライマーは既知の断片の5'伸長を開始するべく合成し、他方のプライマーは既知の断片の3'伸長を開始するべく合成した。開始プライマーの設計は、長さが約22~30ヌクレオチド、GC含有率が50%以上となり、約68~72の温度で標的配列にアニーリングするように、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伸長は全て回避した。

【0272】

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。

【0273】

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200 nmolの各プライマー、 Mg^{2+} と $(NH_4)_2SO_4$ と β -メルカプトエタノールを含むバッファー、Taq DNAポリメラーゼ(Amersham Pharmacia Biotech)、ELONGASE酵素(Life Technologies)、Pfu DNAポリメラーゼ(Stratagene)を含む。プライマーの組、PCI AとPCI Bに対して以下のパラメータで増幅を行った。ステップ1: 94, 3分、ステップ2: 94, 15秒、ステップ3: 60, 1分、ステップ4: 68, 2分、ステップ5: ステップ2、3および4を20回反復する。ステップ6: 68, 5分、ステップ7: 4で保存する。別法では、プライマーの組、T7とSK+に対して以下のパラメータで増幅を行った。ステップ1: 94, 3分、ステップ2: 94, 15秒、ステップ3: 57, 1分、ステップ4: 68, 2分、ステップ5: ステップ2、3および4を20回反復する。ステップ6: 68, 5分、ステップ7: 4で保存する。

【0274】

各ウェルのDNA濃度は、1X TE及び0.5 μ lの希釈していないPCR産物に溶解した100 μ lのPICOGREEN定量試薬(0.25 (v/v) PICOGREEN; Molecular Probes, Eugene OR)を不透明な蛍光光度計プレート(Coming Costar, Acton MA)の各ウェルに分配してDNAが試薬と結合できるようにして測定する。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5~10 μ lを1%アガロースミニゲル上で電気泳動法によって解析し、どの反応が配列の伸長に成功したかを決定した。

【0275】

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レアーゼ(Molecular Biology Research, Madison WI)を用いて消化し、pUC18ベクター(Amersham P

10

20

30

40

50

harmacia 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・シーケンシングのために、消化したヌクレオチドを低濃度(0.6~0.8%)のアガロースゲル上で分離し、断片を切除し、寒天をAgar ACE (Promega)で消化した。伸長させたクローンをT4リガーゼ(New England Biolabs, Beverly MA)を用いてpUC 18ベクター(Amersham Pharmacia Biotech)に再連結し、Pfu DNAポリメラーゼ(Stratagene)で処理して制限部位の張出部(overhang)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LB/2Xカルベニシリン培養液の384ウェルプレートにて37で一晚培養した。

【0276】

細胞を溶解して、Taq DNAポリメラーゼ(Amersham Pharmacia Biotech)及びPfu DNAポリメラーゼ(Stratagene)を用いて以下の手順でDNAをPCR増幅した。ステップ1: 94, 3分、ステップ2: 94, 15秒、ステップ3: 60, 1分、ステップ4: 72, 2分、ステップ5: ステップ2、3および4を29回反復する。ステップ6: 72, 5分、ステップ7: 4で保存する。上記したようにPICOGREEN試薬(Molecular Probes)でDNAを定量した。DNAの回収率が低いサンプルは、上記と同一の条件を用いて再増幅した。サンプルは20%ジメチルスルホキシド(1:2, v/v)で希釈し、DYENAMIC 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DYENAMIC DIRECT kit(Amersham Pharmacia Biotech)またはABI PRISM BIGDYE ターミネーターサイクル シークエンシング反応キット(Terminator cycle sequencing ready reaction kit)(Applied Biosystems)を用いてシークエンシングした。

【0277】

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。

【0278】

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

SEQ ID NO: 20-38から導き出されたハイブリダイゼーションプローブを用いて、cDNA、mRNA、またはゲノムDNAをスクリーニングする。約20塩基対からなるオリゴヌクレオチドの標識について特に記載するが、より大きなヌクレオチド断片に対しても事実上同一の手順が用いられる。オリゴヌクレオチドは、OLIGO 4.06ソフトウェア(National Biosciences)等の最新ソフトウェアを用いて設計し、各オリゴマー50pmolと、[γ -³²P]アデノシン3リン酸(Amersham Pharmacia Biotech)250µCiと、T4ポリヌクレオチドキナーゼ(DuPont NEN, Boston MA)を混合することにより標識する。標識したオリゴヌクレオチドは、SEPHADEX G-25超細繊維分子サイズ排除デキストラン ビードカラム(Amersham Pharmacia Biotech)を用いて十分に精製する。Ase I、Bgl II、Eco RI、Pst I、Xba IまたはPvu II(DuPont NEN)のいずれか1つのエンドヌクレアーゼで消化されたヒトゲノムDNAの典型的な膜ベースのハイブリダイゼーション解析において、毎分10⁷カウントの標識されたプローブを含むアリコットを用いる。

【0279】

各消化物から得たDNAは、0.7%アガロースゲル上で分画してナイロン膜(Nytran Plus, Schleicher & Schuell, Durham NH)に移す。ハイブリダイゼーションは、40で16時間行う。非特異的シグナルを除去するため、例えば0.1xクエン酸ナトリウム食塩水及び0.5%ドデシル硫酸ナトリウムに

10

20

30

40

50

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。

【0280】

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Bal des chweiler等を参照）、機械的マイクロスポッティング技術及びこれらから派生したものをを用いて達成することが可能である。上記各技術において基質は、均一且つ非多孔性の固体とするべきである（S ch e n 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、手作業で、または利用可能な方法や機械を用いて作製でき、任意の適正数のエレメントを有し得る（S c h e n a , M . ら (1995) Science 270:467-470、Sh a l l o n . D . ら (1996) Genome Res. 6:639-645、M a r s h a l l , A . および J . H o d g s o n (1998) Nat. Biotech nol. 16:27-31.を参照）。

10

【0281】

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。

20

30

【0282】

組織または細胞サンプルの調製

グアニジウムチオシアネート法を用いて組織サンプルから全RNAを単離し、オリゴ（dT）セルロース法を用いてポリ（A）⁺ RNAを精製する。各ポリ（A）⁺ RNAサンプルは、MMLV逆転写酵素、0.05pg/μlのオリゴ（dT）プライマー（21mer）、1×第1鎖バッファ、0.03unit/μlのRNAアーゼ阻害剤、500μMのdATP、500μMのdGTP、500μMのdTTP、40μMのdCTP、40μM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。各反応サンプル（1つはCy3、もう1つはCy5標識）は、2.5mlの0.5M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、1mlのグリコーゲン（1mg/ml）を用いて析出させたエタノール、60mlの酢酸ナトリウム及び300mlの100%エタノールである。サンプルは次に、Speed VAC（Savant Instruments Inc., Holbrook N

40

50

Y)を用いて乾燥して仕上げ、 $14\ \mu\text{l}$ の $5\times\text{SSC}/0.2\%\text{SDS}$ 中で再懸濁する。

【0283】

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 $1\sim 2\ \text{ng}$ の初期量から $5\ \mu\text{g}$ より大きい最終量までアレイエレメントを増幅する。増幅されたアレイエレメントは、SEPHACRYL-400 (Amersham Pharmacia Biotech)を用いて精製される。

【0284】

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、蒸留水中で広範囲にわたって洗浄し、 95% エタノール中で 0.05% アミノプロピルシラン(Sigma)を用いてコーティングする。コーティングしたスライドガラスは、 110°C のオープンで硬化させる。

【0285】

米国特許第5,807,522号に記載されている方法を用いて、コーティングしたガラス基板にアレイエレメントを付加する。この特許に引用することを以って本明細書の一部とする。平均濃度が $100\ \text{ng}/\mu\text{l}$ のアレイエレメントDNA $1\ \mu\text{l}$ を高速ロボット装置により開口キャピラリープリントエレメントに充填する。装置はここで、スライド毎に約 $5\ \text{nL}$ のアレイエレメントサンプルを加える。

【0286】

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

【0287】

ハイブリダイゼーション

ハイブリダイゼーション反応は、 $5\times\text{SSC}$, $0.2\%\text{SDS}$ ハイブリダイゼーション緩衝液中のCy3及びCy5標識したcDNA合成生成物を各 $0.2\ \mu\text{g}$ 含む $9\ \mu\text{l}$ のサンプル混合体を有する。サンプル混合体は、 65°C まで5分間加熱し、マイクロアレイ表面上で等分して $1.8\ \text{cm}^2$ のカバーガラスで覆う。アレイは、顕微鏡スライドより僅かに大きい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 $140\ \mu\text{l}$ の $5\times\text{SSC}$ を加えることにより、チェンバー内部を湿度 100% に保持する。アレイを含むチェンバーは、 60°C で約6.5時間インキュベートする。アレイは、第1洗浄緩衝液中($1\times\text{SSC}$, $0.1\%\text{SDS}$)において 45°C で10分間洗浄し、第2洗浄緩衝液中($0.1\times\text{SSC}$)において 45°C で10分間各々3度洗浄して乾燥させる。

【0288】

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 $488\ \text{nm}$ 、Cy5の励起のためには $632\ \text{nm}$ でスペクトル線が発生し得るInnova 70混合ガス $10\ \text{W}$ レーザ(Coherent, Inc., Santa Clara CA)を備えた顕微鏡で検出する。 $20\times$ 顕微鏡対物レンズ(Nikon, Inc., Melville NY)を用いて、アレイ上に励起レーザ光の焦点を当てる。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

10

20

30

40

50

ースキャンする。本実施例で用いた $1.8\text{ cm} \times 1.8\text{ cm}$ のアレイは、 $20\text{ }\mu\text{m}$ の解像度でスキャンした。

【0289】

2つの異なるスキャンのうち、混合ガスマルチラインレーザは2つの蛍光色素を連続的に励起する。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

【0290】

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 (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って行う。

20

【0291】

光電子増倍管の出力は、IBM 互換 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

【0292】

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

【0293】

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 $15 \sim 30$ 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

【0294】

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、抗生物質耐性遺伝子及び 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

50

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 *Autographica californica* 核多面性ウイルス(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可欠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関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)。

10

【0295】

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。

20

30

【0296】

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。選択ベクターには、pCMVSPORT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

40

50

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。

【0297】

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。

10

【0298】

1.4 GCREC特異抗体の産生

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。

20

【0299】

或いは、レーザGENEソフトウェア(DNA STAR)を用いてGCRECアミノ酸配列を解析し、免疫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

【0300】

通常は、長さ約15残基のオリゴペプチドを、Fmocケミストリを用いるABI 431A ペプチドシンセサイザ(Applied Biosystems)を用いて合成し、N-マレイミドベンゾイル-N-ヒドロキシスクシンイミドエステル(MBS)を用いた反応によってKLH(Sigma-Aldrich, St. Louis MO)に結合させて、免疫原性を高める(前出のAusubel, 1995等を参照)。完全フロイントアダ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

【0301】

1.5 特異抗体を用いた天然の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(Amersham 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。

【0302】

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

50

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ップ剤)でカラムを溶出させ、GCRECを回収する。

【0303】

16 GCRECと相互作用する分子の同定

GCRECまたは生物学的に活性であるGCREC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. and W. M. Hunter (1973) *Biochem. J.* 133:529-539を参照。)マルチウエ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。様々なGCREC濃度で得られたデータを用いて、候補分子と結合したGCRECの数量及び親和性、会合についての値を計算する。

10

【0304】

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。

【0305】

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

20

17 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系に形質移入する。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。

30

【0306】

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の密度として定義され、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頁)。

40

【0307】

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,

50

P. ら (1998) *J. Biol. Chem.* 273:4990-4996) を調整する能力に基づく。当分野で既知の方法を用いて、完全長 GCR E C をコードするプラスミドを哺乳動物細胞系 (例えばチャイニーズハムスター卵巣 (CHO) またはヒト胎児腎臓 (HEK-293) 細胞系) に形質移入する。培地中の 12 穴トレイで形質移入された細胞を成長させ、培養液を廃棄して、接着した細胞を PBS で軽く洗浄する。リガンドを入れた培養液とリガンドを入れない培養液で中で細胞を 30 分間インキュベートし、次に培養液を除去し、細胞を 1 M 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 レベルをリガンドを入れなかった場合と比較したときの变化は、形質移入された細胞に存在する GCR E C の量に比例する。

10

【0308】

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養液及び [^3H] 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。培養液を除去し、10 mM 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变化は、形質移入された細胞に存在する GCR E C の量に比例する。

20

【0309】

1.8 GCR E C リガンドの同定

GCR E C 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 などの真核細胞系において発現する。これらの細胞系は、GPCR 発現過程が良好であり、発現した GCR E C 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+} などの細胞内セカンド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、これらセカンド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 Fluor-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セカンドメッセンジャー経路が知られていない場合、GCR E C 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 GCR E C のシグナル伝達を通すために、広範囲の G タンパク質と共役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 (Offermanns, S. and M. I. Simon (1995) *J. Biol. Chem.* 270:15175-15180)。或いは、内在性 GPCR が不足しているように遺伝子操作された酵母系に GCR E C を発現し、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、内在性酵母フェロモン受容体経路の対応する構成要素をヒト GPCR 及び G タンパク質で置換する。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. and J. Thorne (1996) *Nature* 384 (supp.): 14-16)。既知の GPCR リガ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

30

40

50

して受容体をスクリーニングする。組織、生物学的液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。

【0310】

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。

【0311】

10

(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。

【0312】

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。

【0313】

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

20

【0314】

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。

【0315】

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。

【0316】

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。

【0317】

30

表7は、NTTの分析に用いたツール、プログラム、アルゴリズムを、適用可能な説明、引用文献及び閾値パラメータと共に示す。

【表1】

表 1

Incyte プロジェクト ID	ポリバプタシド SEQ ID NO:	Incyte ポリバプタシド ID	ポリスクレオノイド SEQ ID NO:	Incyte ポリスクレオノイド ID
7474806	1	7474806CD1	20	7474806CB1
7474840	2	7474840CD1	21	7474840CB1
7475092	3	7475092CD1	22	7475092CB1
7341260	4	7341260CD1	23	7341260CB1
7473911	5	7473911CD1	24	7473911CB1
7474767	6	7474767CD1	25	7474767CB1
7475815	7	7475815CD1	26	7475815CB1
60263275	8	60263275CD1	27	60263275CB1
60203310	9	60203310CD1	28	60203310CB1
7477349	10	7477349CD1	29	7477349CB1
55002235	11	55002235CD1	30	55002235CB1
7475686	12	7475686CD1	31	7475686CB1
7482007	13	7482007CD1	32	7482007CB1
6769042	14	6769042CD1	33	6769042CB1
7476053	15	7476053CD1	34	7476053CB1
7480410	16	7480410CD1	35	7480410CB1
55036418	17	55036418CD1	36	55036418CB1
7481701	18	7481701CD1	37	7481701CB1
7481774	19	7481774CD1	38	7481774CB1

10

20

30

40

【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:	標準スコア	GenBank 相同体
1	7474806CD1	g2707256	4.00E-62	[f1][シチメンチョウ (Meleagris gallopavo)] Gタンパク質共役2Yスクレオチド受容体
2	7474840CD1	g11967419	4.00E-65	[f1][マウス]鋸骨器官(プロモモン)受容体 VIRC3
3	7475092CD1	g3992628	2.00E-86	[f1][ヒト] 推定7パス膜貫通タンパク質
4	7341260CD1	g3805932	1.00E-10	[f1][ヒト] 推定Gタンパク質共役受容体, EDG6 Graier, M.H.他 (1998) Genomics 53:164-169
5	7473911CD1	g1055254	2.00E-18	[f1][ドブネズミ] プロモモン受容体 VN6 Dulac, C. and Axel, R. (1995) Cell 83:195-206
6	7474767CD1	g179985	2.00E-14	[f1][ヒト] C-C ケモカイン受容体 1型 Neote, K. 他 (1993) Cell 72:415-425
7	7475815CD1	g1055254	1.00E-56	[f1][ドブネズミ] プロモモン受容体 VN6
8	60263275CD1	g13183149	0	[f1][ヒト] (AF239764) EGF様モジュラ含有ムチン様受容体 BMR3
9	60203310CD1	g3882981	0	[f1][ドブネズミ] カルジウム依存性アルブアラトロトキシン受容体
10	7477349CD1	g6979162	3.00E-11	[f1][ドブネズミ] マクロブアラージ炎症性タンパク質-1アルファ受容体
11	55002225CD1	g14164383	0	[f1][ヒト] (AB060151) Gタンパク質共役受容体
12	7475686CD1	g7248884	1.00E-180	[f1][マウス] G-タンパク質共役受容体 GPR73 Parker, R. 他(2000) Biochim. Biophys. Acta 1491:369-375
13	7482007CD1	g5525078	1.00E-104	[f1][ドブネズミ] 膜7回貫通受容体 Abe, J. 他 (1999) J. Biol. Chem. 274:19957-19964.
14	6769042CD1	g4164061	1.70E-59	[ウツ] ラトロフィリン3 スブタイプ変異体 abbg FEBS Lett. (1999) 443:348-352
15	7476053CD1	g330075	1.00E-165	[f1][ドブネズミ] セトロニン受容体 Briander, M.G. et al. (1993) Proc. Natl. Acad. Sci. U.S.A. 90:3452-3456
16	7480410CD1	g39983382	1.90E-88	[ハツカネズミ] 嗅覚受容体 E3

10

20

30

40

表2-2

ポリペプチド SEQ ID NO:	Incyte ポリペプチド ID	GenBank ID NO:	継続スコア	GenBank 相同体
17	55036416CD1	G3983398	5.00E-99	[ハツカネズミ] 嗅覚受容体 G3 Krautwurst, D. 他 (1998) Cell 95:917-926
18	7481701CD1	g12007416	2.00E-67	[f1] [マウス] m51 嗅覚受容体
19	7481774CD1	95901478	3.10E-106	[マウス] 嗅覚受容体

10

20

30

40

表3-1

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
1	7474806CD1	339	S111 S15 S189 S260 S336 S46 S84 T163 T204 T255 T36 T80	N35 N69	膜貫通ドメイン: Y89-T109, F130-F149 膜7回貫通受容体 (ロドプシンファミリー): L68-Y324 Gタンパク質共役受容体 BL00237: M116-P155, P224-Y235, T255-T281, S316-I332 ロドプシン様GPCRスーパーファミリー-PR00237: Y53-W77, A86-I107, F130-V152, H166-A187, G216-V239, S260-F284, Y306-I332 Gタンパク質共役受容体: DM00013 P4123 I27-322: F45-S327 P51582 29-322: F45-S327 I55450 20-317: F45-R328 P48042 45-340: Y53-S327 Gタンパク質共役受容体: PD000009: Y89-F188 フェロモン受容体: PD009900: I59-M329	HMMER HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-DMC BLAST-PRODOM BLAST-PRODOM
2	7474840CD1	335	S237 S28 S318 T22 T86	N168 N184	膜貫通ドメイン: H166-L186, M240-Y256, Y299-L319 推定7回パス膜貫通タンパク質: ED138976: R3-L360	HMMER BLAST-PRODOM
3	7475092CD1	428	S257 S286 S295 S342 S356 S369 S412 S79 T156	N180 N207 N284	膜貫通ドメイン: L55-L74 膜7回貫通受容体 (ロドプシンファミリー): A31-A250	HMMER-PFAM BLIMPS-PRINTS
4	7341260CD1	330	S123 S195 S323 S5 S78 T219	N154 N4 N76 N93	ロドプシン様GPCRスーパーファミリー: PR00237B:A50-L71 PR00237C:P92-V114 PR00237D:S123-L144 PR00237F:Q222-L246 PR00237G:L262-T288	HMMER BLIMPS-PRINTS
5	7473911CD1	676	S105 S213 S215 S250 S330 S353 S492 S569 S591 S606 S646 S647 S79 T217 T296 T301 T394 T620	N184	膜貫通ドメイン: I118-A136 フェロモン受容体 PD009900: S87-V167, K42-Y90	HMMER BLAST-PRODOM
6	7474767CD1	372	S214 S289 S351 T145 T223 T299	N11 N318 N353	シグナルペプチド: M1-T47 膜貫通ドメイン: Y32-SE2, I74-P100, I184-Y208	SPSCAN HMMER

10

20

30

40

【表5】

表3-2

SEQ ID NO.	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
7	7475815CD1	271	S115 S142 S188 S203 S79 T269		膜7回貫通受容体 (ロドプシンファミリー) : A43-Y285 G-タンパク質共役受容体タンパク質 BL00237:M94-P133, Y221-N247, N277-R293 ロドプシン様GPCRスーパーファミリーシグネチャ: PR00237: V28-S52, S62-V83, E108-V130, R144-W165, I184-V207, T226-Y250, S267-R293 G-タンパク質共役受容体シグネチャ: E105-T155 C-タンパク質共役受容体: DM00013: P22246 28-316:A22-R293 P51677 28-316:F27-R293 P25930 33-319:F27-L300 P34981 19-336:V28-I206, S222-R293 膜貫通ドメイン: I216-A234 フェロモン受容体 PD009900 : K42-S170, I112-K271 シグナルペプチド: M1-Q22 膜貫通ドメイン: V321-I339, W424-T451, M464-I488, Y540-L560 膜7回貫通受容体 (セクレチンファミリー) : D312-V564 Letrophilin/CL-I様 GPCR ドメイン: K259-Q309 EGF様ドメイン C30-C76	HMMER_PPFAM BLIMPS_BLOCKS BLIMPS_PRINTS PROFILESSCAN BLAST_DCMO HMMER BLAST_PRODOM
8	60263275CD1	611	S257 S308 S494 S578 S580 S588 S598 S63 S81 T21 T211 T258 T307 T406 T75 Y885	N104 N148 N161 N209 N238 N286 N293 N345 N409 N414	シグナルペプチド: M1-Q22 膜貫通ドメイン: V321-I339, W424-T451, M464-I488, Y540-L560 膜7回貫通受容体 (セクレチンファミリー) : D312-V564 Letrophilin/CL-I様 GPCR ドメイン: K259-Q309 EGF様ドメイン C30-C76 G-タンパク質共役受容体 BL00649: C378-L403, G430-R454, W465-S494 セクレチン様GPCRスーパーファミリーシグネチャ PR00249: V317-K341, I380-L403, K423-L448 W465-K490, A539-L560 I型EGFシグネチャ PR00009 : K25-F40, E48-Y59 G-タンパク質共役受容体ファミリー 2 シグネチャ: V531-E574	HMMER HMMER HMMER_PPFAM HMMER_PPFAM BLIMPS_BLOCKS BLIMPS_PRINTS BLIMPS_PRINTS BLIMPS_PRINTS PROFILESSCAN

10

20

30

40

【表6】

表3-3

SEQ ID NO.	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシ化部位	シグネチン配列、ドメインおよびモチーフ	分析方法及びデータベース
9	6C203310C21	1469	S166 S186 S1214 N464 N549 S187 S261 T1253 N759 N772 S225 S351 S1295 N817 N843 N93 S417 S423 T1398 N932 N1098 S460 S591 T1453 N1179 N1196 S630 S761 T1008 N1251 S791 S824 S1144 S868 S907 S1154 S978 T113 T1181 T181 T199 S1295 T294 T303 T1346 T443 T559 S1458 T658 T672 T731 T803 T842 Y121		シグナルペプチド: M1-S21 膜貫通ドメイン: V879-F897, V942-L960, F1020-G1040, E1071-W1090 膜7回貫通受容体 (セクレチンファミリー): D874-V1130 Lactrophilin/CL-1様 GPS ドメイン: F814-V866 Gタンパク質共役受容体 BL00649: C940-L965, G987-V1011, W1022-A1051, S1113-L1138, C498-T525, G884-I929 セクレチン様 GPCR スーパーファミリー PR00249: V879-R903, V942-L965, R980-S1005, W1022-V1047, A1105-L1126 EMR1 ホルモン白血球抗原: DM05221 I37225 347-738: L769-S1160 P48960 347-738: L769-S1160 A57172 465-886: N730-T1162 Gタンパク質共役受容体ファミリー 2: DM00378 I49149 56-425: G836-R1131 Lactrophilin 関連受容体: PD024331: K386-E628 PD041747: L1207-L1469	BLAST_PRODOM BLAST_PRODOM BLAST_DOMO BLAST_DOMO MOTIFS HMMER HMMER HMMER-PFAM HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-DOMO BLAST-PRODOM

10

20

30

40

【表7】

表3-4

SEQ ID NO:	Incyte ポリヌクレオチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的リン酸化部位潜在的グリコシ化部位	シグネチャ配列、ドメインおよびモジュール	分析方法及びデータベース
10	7477349CD1	469	S122 S199 S25 S312 S447 S85 T110 T195 T207 T282 T315 T39 T427 T431 T465	N139	ミオソリン/オルファクトメドイン (Olfactomedin) 応答タンパク質: EGF様Gタンパク質共役受容体: ED006897: Y177-V385 EGF様Gタンパク質共役受容体: ED005428: R629-V866 縦貫ドメイン: P166-L188, V215-V233 膜7回貫通受容体 (ロドプシンファミリー): A251-Y421 Gタンパク質共役受容体モジュール: Gタンパク質共役受容体シグネチャ: I243-A293 Gタンパク質共役受容体 BL00237: R231-P270, Y329-R340, P357-V363, H413-R429 ロドプシン様GPCRスーパーファミリー: I165-A189, S199-A220, E245-L267, R281-W302, K321-I344, S362-Y386, L403-R429 Gタンパク質共役受容体: DM00013 I49339 28-316: A162-V435 P51675 28-317: A162-V435	BLAST-PRODOM BLAST-PRODOM HMMER HMMER-PTAM MOTIFS ProfileScan ELIMPS-BLOCKS ELIMPS-PRINTS ELAST-DOMO ELAST-DOMO

10

20

30

40

表3-5

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
11	55902225CD1	335	S190 S30 S62 T11 T110 T150 T189 T320 T330 T66	N10 N17	G タンパク質共役受容体 DM00011:P32745 38-329;V32-S310 受容体共役 G タンパク質 膜貫通 糖蛋白質リン酸化 リポ蛋白質 ハルミチン酸 タンパク質 アミノ酸 PD000009:R61-Y171 G タンパク質共役受容体 BL00237:S298-Q314, W99-P138, F206-Y217, Q244-I270 ロドプシン様 GPCR スーパーファミリーシグネチャ PR00237:I36-I60, P68-L89, D113-L135, K149-V170, T198-L221, L249-V273, Y288-Q314 膜貫通ドメイン transmem_domain:T34-F57, V194-I218, M252-I272 膜7 同貫通受容体 (ロドプシンファミリー) 7Lm_1:G51-Y306 G Protein Receptor A119-L135 G タンパク質共役受容体シグネチャ g_protein_receptor.prf:T110-L157	BLAST_DOMC BLAST_PRODOM BLIMPS_BLOCKS BLIMPS_PRINTS HMMER HMMER_PFBM MOTIFS PROFILESAN
12	7475666CD1	630	S230 S269 S601 S607 T223 T519 T581 T586 T620 T81	N332 N584	G タンパク質共役受容体 DM00013:P49146 44-341:K299-F590 受容体共役 G タンパク質 膜貫通 糖蛋白質リン酸化 リポ蛋白質 ハルミチン酸 タンパク質 アミノ酸 PD000009:K328-F437 G タンパク質共役受容体 BL00237:W367-F406, F478-Y489, R514-F540 ロドプシン様 GPCR スーパーファミリーシグネチャ PR00237:R381-I403, T415-Y436, F470-S493, T519-V543, F561-M587, V301-T325, T334-F355 神経ペプチド Y 受容体 PR01012:R326-I338, R381-A396 膜貫通ドメイン transmem_domain: I302-F320	BLAST_DOMC BLAST_PRODOM BLIMPS_BLOCKS BLIMPS_PRINTS HMMER

10

20

30

40

【表 9】

表3-6

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
13	7482007CD1	635	S185 S252 S354 S392 S395 S591 S609 I265 T397 T45 T584 T624 T653	N169 N177 N209 N229 N250 N257 N263 N296 N309 N340 N379 N61 N679 N686	膜7回貫通受容体 (ロドプシンファミリー) 7tm_1:G316-N575 G Protein Receptor V387-I403 Gタンパク質共役受容体シグネチャ g_protein_receptor.prf:N378-I428 ホルモン; EMRI; 白血球; 抗原; DM05221 I37225 347-738:C349-S688 受容体 膜貫通 Gタンパク質共役 糖蛋白 前駆体 シグナルタイプ ポリペプチド代替 PD000752:N372-R660 Gタンパク質共役受容体ファミリー 2 タンパク質 BL00649:C473-I498 セクレチン様 GPCR スーパーファミリーシグネチャ PR00249:Y403-W427, A475-I498 膜貫通ドメイン transmem_domain:I407-T433, M514-T537, I487-F506, I559-V579 膜7回貫通受容体 (セクレチンファミリー) 7tm_2:D398-I659	HMMER_Pfam MOTIFS PROFILESKAN BLAST_DOMC BLAST_PRODOM BLIMPS_BLOCKS BLIMPS_PRINTS HMMER HMMER_Pfam
14	6769042CD1	633	S108 S262 S275 S381 S589 S612 S616 S70 T251 T355 T517	N153 N235 N260 N263 N292 N366 N41-N604 N61 N78	ホルモン; EMRI; 白血球; 抗原; DM05221 A57172 465-886:G279-S593 受容体 膜貫通 Gタンパク質共役 糖蛋白 前駆体 シグナルタイプ ポリペプチド代替 PD000752:H321-K579 Gタンパク質共役受容体ファミリー 2 タンパク質 BL00649:S555-I580, C391-L416 GAMP-タイプ GPCR シグネチャ PR00247:Y361-R383, A395-I421, Y433-S451 セクレチン様 GPCR スーパーファミリーシグネチャ PR00249:S327-S351, V393-I416, H431-S456, W473-S498, T517-I537, Q547-I568 膜貫通ドメイン transmem_domain: C333-L350, I472-T494 膜7回貫通受容体 (セクレチンファミリー) 7tm_2:Q322-V572	BLAST_DOMC BLAST_PRODOM BLIMPS_BLOCKS BLIMPS_PRINTS BLIMPS_PRINTS HMMER HMMER_Pfam

10

20

30

40

表3-7

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法およびデータベース
15	7476053CD1	370	S217 S27 S287 T116 T163		Latrophilin/CL-1様 GFS ドメイン: GPS: V265-L315 Gタンパク質共役受容体ファミリー2 シグネチャ g_protein_recep_f2_2.prf: A474- S593 シグナルペプチド: M298-T317 Signal_cleavage: M1-A65 膜貫通ドメイン: F48-L68, M298-L316 膜7回貫通受容体 (ロドプシンファミリー) 7tm_1: M69-Y351 Gタンパク質共役受容体 BL00237: R120-H159, Q291-T317, N343-N359 ロドプシン様 GPCR スーパーファミリー シグネチャ PR00237: L54-P78, P87-F108, D134-I156, A170-L191, A214-Y237, A296-I320, K333-N359 5-ヒドロキシトリプタミン 5B 受容体 PR00519: A3-R19, E20-P36, P36-V50, E196-R204, R246-V260, V260-V270 5ヒドロキシトリプタミン 5B 受容体 5HT5B セロトニン Gタンパク質共役 膜貫通 糖タンパク質 多重遺伝子 PD027821: M1-H84 Gタンパク質共役受容体 DM00013: P31387 46-367: P46-T367 I48231 46-367: P46-T367 P47898: 34-354: P46-T367 P20905: I56-522: P46-S280 シグナルペプチド: M1-A39 膜貫通ドメイン: G26-I50, C203-V217 膜7回貫通受容体 (ロドプシンファミリー): A42-Y291 Gタンパク質共役受容体シグネチャ: F103-V147 Gタンパク質共役受容体モチーフ: G111-V127 Gタンパク質共役受容体 BL00237: T283-A299, K91-P130, L208-Y219, R236-L262	HMMER_PPFAM PROFILES SCAN HMMER SPSCAN HMMER HMMER_PPFAM BLIMPS_BLOCKS BLIMPS_PRINTS BLIMPS_PRINTS BLAST_PRODOM BLAST_DOMO SPSCAN HMMER HMMER_PPFAM ProfileScan MOTIFS BLIMPS_BLOCKS
16	7480410CD1	324	S189 S194 S292 S314 S58	N9		SPSCAN HMMER HMMER_PPFAM ProfileScan MOTIFS BLIMPS_BLOCKS

10

20

30

40

【表 1 1】

表3-8

SEQ ID NO.	Incyte ポリペプチドID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
17	55036418CD1	315	S194 S22 S292 S68 S8 S88	N6	ロドプシン様 GPCR スーパーファミリー- PR00237: L27-H51, M60-K81, Y105-V127, M200-L233, P25-L49, K273-A299 嗅覚受容体シグネチャ PR00245: M60-K81, F178-D192, F239-G254, V275-L286, S292-L306 Gタンパク質共役受容体: DM00013 P23275 17-306: H21-G307 A57069 15-304: H21-G307 P30954 29-316: L28-G307 S29709 11-299: P25-G307 嗅覚Gタンパク質共役受容体: PD000921: F169-N247 嗅覚Gタンパク質共役受容体: PD149621: V248-R308 シグナルペプチド: M1-A24 膜貫通ドメイン: L26-L46, C98-M119, F201-A220 膜7回貫通受容体(ロドプシンファミリー): G42-Y291 Gタンパク質共役受容体モチーフ: S111-I127 Gタンパク質共役受容体シグネチャ: L104-G147 BL00237: K91-P130, L208-Y219, K236-R262, T283-M299 ロドプシン様 GPCR スーパーファミリー PR00237: V27-Y51, M60-K81, F105-I127, I200-L233, K273-M299 嗅覚受容体シグネチャ PR00245: M60-K81, F178-D192, L239-G254, A275-L286, S292-L306 I18-L302 A57069 15-304: F19-K304 P34982 17-305: V27-D307 P23270 18-311: L26-L306	BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-DOMO BLAST-PRODOM SPScan HMMER HMMER-PFAM MOTIFS ProfileScat BLIMPS-BLOCKS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-DOMO

10

20

30

40

表3-9

SEQ ID NO.	Incyte 飛バ ブナドID	アミノ酸 残基数	潜在的リン酸化部位	潜在的グリコシ 化部位	シグネチャ配列、 ドメインおよびモチーフ	分析方法及びデ ータベース
18	7481701CD1	324	S292 S314 S320 S67 S87 T136 T8		喫覚Gタンパク質共役受容体 ; PD000921: Y169-I1246 喫覚Gタンパク質共役受容体 ; PD149621: T247-R308 膜7回貫通受容体 (ロドプシンファミリー) : G41- F180, V225-Y291 視覚色素 (オプシン) 網膜結合部: W271-T319 Gタンパク質共役受容体 BL00237: R89-F128, R236-R262, S283-K299 ロドプシン様GPCR スーパーファミリー PR00237: V26-Q50, M59-K80, F103-I125, F12-F33, T144-F167, A238-R262, I273-K299 喫覚受容体シグネチャ PR00245: M59-K80, I176-D190, F239-G254, S292-I306 メラノコルチン受容体ファミリー PR00534: L51-I63, I125-T136 Gタンパク質共役受容体 : DM00013 P23267 20-309; F17-I306 S29709: I1-299; V29-G307 P23270 18-311; F17-K304 P23274 18-306; V26-I302 喫覚Gタンパク質共役受容体 : PD000921: L165-I247 膜貫通ドメイン: I24-V46, Q98-M116, F198-L214 膜7回貫通受容体 (ロドプシンファミリー) : G39-Y288 Gタンパク質共役受容体モチーフ: V108-I124 F100-T146 Gタンパク質共役受容体 BL00237: K88-F127, V205-Y216, Q233-Q259, T280-K296 喫覚受容体シグネチャ PR00245: M57-Q78, F175-S189, F236-G261, L272-I283, T289-L303	BLAST-PRODOM BLAST-PRODOM HMMER HMMER-PFAM ProfileScan BLIMPS-BLOCKS BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-DMO
19	7481774CD1	312	S16 S230 S264 S47 S65 T161 T289	N40	喫覚Gタンパク質共役受容体 ; PD000921: L165-I247 膜貫通ドメイン: I24-V46, Q98-M116, F198-L214 膜7回貫通受容体 (ロドプシンファミリー) : G39-Y288 Gタンパク質共役受容体モチーフ: V108-I124 F100-T146 Gタンパク質共役受容体 BL00237: K88-F127, V205-Y216, Q233-Q259, T280-K296 喫覚受容体シグネチャ PR00245: M57-Q78, F175-S189, F236-G261, L272-I283, T289-L303	BLAST-PRODOM HMMER HMMER-PFAM MOTIFS ProfileScan BLIMPS-BLOCKS BLIMPS-PRINTS

10

20

30

40

表3-10

SEQ ID NO:	Incyte 非リベ ブチドID	アミノ酸 残基数	潜在的リン酸化部位 潜在的グリコシ ル化部位	シグネチャ配列、 ドメインおよびモチーフ	分析方法及び データベース
				Gタンパク質共役受容体 : DM00013 P23275 17-305: S16-L303 A57069 15-304: F15-L303 P34982 17-305: S16-L303 P30953 18-306: R18-L303 膜貫通タンパク質共役受容体 : PD149521: T244-R307 膜貫通タンパク質共役受容体 : PD000921: C157-L243	BLAST-DMO BLAST-PRODOM BLAST-PRODOM

10

20

30

40

表4-1

ポリスクレオチド SEQ ID NO.	Incyte クローンID	ポリ 配列長	選択された断片	配列断片	5'位置	3'位置
20	7474806CB1	1076	1-1076	7075196HI (BRAUTDR04) 97248967 edit 97407927_edit	1 533 1	532 1076 1102
21	7474840CB1	1102	1007-1102, 107-160, 486-522, 642-842	55049853HI 71906055V1 1351856F6 (LATRUT02) GNN.97023955_000031_002 71900320V1 8017335J1 (BMARTX01) 55049805J1 7341260HI (COLNDIN02) 70811587V1 6834479HI (BRSTNON02) 70888652V1 7628842HI (GLADIE01) FL140044_00001	876 564 1757 1 1988 64 3256 307 991 974 228 660 1 1	1612 1209 2367 468 2529 614 1956 1847 1575 284 2031
22	7475092CB1	2529	628-1475, 1-81	GNN.96693326_000106_002 55093139J1	139	1130
23	7341260CB1	1847	1-140, 1490-1847	CpG_991027_B15_masked_fa. FL7475815_g8492585_000004_ 9392596 71704087V1 71551942V1	1 372 85 986	276 723 1202 700 1727
24	7473911CB1	2031	1-504, 669-1834	3642425T6 (LUNGN0T34) 524802R6 (CARCTXT01) 2435123HI (BRAVUNT02) 7155160V1 71959831V1	1469 391 1 838 4179	2079 985 255 1544 5068
25	7474767CB1	1130	1-1130	7638002J1 (SEMUTDE01) 491493HI (HNT2AGT01)	2283 3833	2900 4073
26	7475815CB1	1202	367-959, 1044-1202, 1-116			
27	60263275CB1	2079	134-490, 902-981			
28	60203310CB1	5324	1-373, 963-3661, 4111-4162			

10

20

30

40

【表 15】

表4-2

ボリヌクレオチド SEQ ID NO:	Incyte ヌクレオチドID	ボリ 配列長	選択された断片	配列断片	5'位置	3'位置				
29	7477349CB1	1962	1-1758	71957696V1	4646	5324				
				4028716F7 (BRAINOI23)	4057	4787				
				8103587H1 (MIXDDIE02)	620	1259				
				9900324	3652	4165				
				8195081H2 (BRAIDIR04)	1	623				
				55073390J1	1549	2191				
				60200671D1	3444	3823				
				8195081J2 (BRAIDIR04)	432	1033				
				55094091J1	1140	2026				
				55116120H1	2658	3416				
30	5500225CB1	1558	80-1558, 1-49	70846216V1	1486	1962				
				GNN_98140731_000028_002	265	1410				
				GNN_99309533_000030_002	1	379				
				2021568F6 (CONNTOI1)	379	873				
				71243436V1	1272	1931				
				70844223V1	872	1416				
				70845761V1	804	1415				
				72398219V1	675	1384				
				72374379V1	971	1558				
				72373094V1	814	1540				
31	7475686CB1	2304	1-1212, 1548-2304	55049494J1	1	817				
				CPG_SAE300482544.R1	1148	1481				
				GNN:96138786_000002_006.ed	1	1893				
				it						
				7290466F6 (BRAIFER06)	1576	2304				
				55049456J1	461	3327				
				6925371H1 (PLACFER06)	249	797				
				GNN:99864547_000007_002_ed	472	2103				
				it						
				55084155J2	1017	1790				
32	7482007CB1	2322	1-628, 1452-1878, 724-1278	7341368F8 (COLNDING02)	1	440				
				91507289	1879	2322				
				7157074H2 (BRAIFEJ02)	1169	1797				
				72138116D1	1704	2366				
				7629227H1 (GELADIE01)	834	1341				
				33	6769042CB1	2366	800-1248, 1-40, 1369-1717,			

10

20

30

40

【表 16】

表4-3

ポリヌクレオチド SEQ ID NO:	Incyte ヌクレオチド ID	ポリ	配列長	選択された断片	配列断片	5'位置	3'位置
				218-361, 1821-1861, 1930-2366	55147608J1	1	908
34	7476053CB1		1458	1-1087	GNN.97630608_000013_022 93280262 GBI:99454621.edit	1201 1088	1414 1458
35	7480410CB1		975	1-816, 894- 975	55036194H1 (GPCRDPV02)	166	371
36	55036418CB1		948	1-162, 193- 948	GNN.98979559_000011_002 55036391H1 (GPCRDPV02)	166	975
37	7481701CB1		1086	1-1086	GNN.97239420_000072_008	1	948
38	7481774CB1		1529	1-963	GNN.99795014_000002_002 55143535J1 70822063V1	127 879	1086 1529
					7361408F8 (BRAIFEE05)	401	1051
					55142634H1	1	384

10

20

30

40

表 5

ボリスケレオサド SEQ ID No:	Incyte プロジェクトID	代表的ライブラリ
20	7474806CB1	BRAUIDR04
22	7475092CB1	LATUTU02
23	7341260CB1	COLNTUT03
24	7473911CB1	BRSTNOT23
27	60263275CB1	EOSITXT01
28	60203310CB1	BRAITUT01
29	7477349CB1	CONNNOT01
31	7475686CB1	BRAIFER06
32	7482007CB1	COLNDIN02
33	6769042CB1	GELADIE01
38	7481774CB1	BRAIFEE05

10

20

30

40

表6-1

ライブラリ	ベクター	ライブラリの説明
BRA1FE05	PCDNA2.1	ライブラリは、妊娠23週間で死産の左心低形成症胎児(白人男子)から除去された脳組織から単離したRNAを用いて作製した。血清検査は陰性だった。
BRA1FE06	PCDNA2.1	このライブラリは50才の白人女性の前置卵巣切除時に除去された脳腫瘍組織から単離したRNAを用いて作製した。血清検査は陰性だった。
BRA1TUT01	PSPORT1	ライブラリは50才の白人女性の前置卵巣切除時に除去された脳腫瘍組織から単離したRNAを用いて作製した。病理は再発性グレード3のオリゴアストロサイトーマを示し、限局性壊死と広汎にわたる石灰化があった。患者の病歴には言語障害と癲癇が含まれる。患者の脳はまた全投与量、5,082 cyg [分画8] の照射を受けた。家族歴には脳腫瘍がある。
BRA1TUT04	PCDNA2.1	このライブラリは55才の白人女性から除去され、プーランドされていた線状体、後尾状核、後極部および嗅球核から単離されたRNAを用いて作製された。病理は弓状部にとつた程度の髄膜線維形成があり、帯状皮質白質および嗅球核に軸索球状体が散在しており、内嗅野および中脳水道灰白質領域にいくぶん神経原線維変化の散在があった。関連腫瘍組織の病理は、残留性または再発性腫瘍のある高分化型の肝臓胆管癌を示した。患者の病歴は胆管癌、術後バッド・キラーリ症候群、胆管癌水腫、水胸症、脱水症、栄養不良、乏尿および急性腎不全が含まれる。前に受けた手術には胆管切除術および肝臓の85%の摘除がある。
BRSTNOT23	pLINC	ライブラリは35才の白人女性の両側性縮小乳房形成術時に採取した乳房組織から単離したRNAを用いて作製した。病理は非増殖性繊維囊胞性疾患を示した。家族歴にはII型糖尿病、アテローム硬化型冠動脈疾患、急性心筋梗塞、高脂血症および冠動脈ハイパスがある。
COLND1N02	pLINC	この標準化したライブラリは、罹患大腸および大腸ポリープ組織ライブラリからの472万個の独立性クローンから作製した。開始RNAは2つの供与者からプーランドされたcDNAから作製した。cDNAは16才白人男子(供与者A)の部分結腸切除、随時回腸人工肛門術および結腸鏡検査時に得た盲腸と下行結腸から除去した罹患大腸組織および67才の女性(供与者B)の盲腸から除去された罹患大腸ポリープ組織から単離されたmRNAを用いて作製した。病理学検査では、家族性多発性ポリープ症での結腸粘膜全体が関与する低グレードの形成異常のある無数の(100以上)腺腫性ポリープ(供与者A)と良性の盲腸ポリープ(供与者B)が見られた。関連する腫瘍組織(B)の病理は、盲腸における固状に発育する塊を形成する管状絨毛状腺腫に生じた浸潤度3の腺癌を示した。腫瘍は固有筋層を越えて浸潤していた。複数の局所リンパ節(17個中2個)が転移性腺癌に関わっている。管状絨毛状腺腫とグレードの低い異形成を伴う複数(6個)の管状腺癌が盲腸および上行結腸に観察された。供与者Aには腫瘍節及び膨満があった。患者が服用している薬剤はなかった。家族歴は父親と兄弟姉妹に良性大腸腫瘍、社父母に良性高血圧、脳血管疾患、乳癌、子宮がんおよびII型糖尿病がある。
COLNUT03	pLINC	ライブラリは67才の白人男性のS状結腸切除術と永久的人工肛門形成術時にS状結腸から得た大腸腫瘍組織から単離したRNAを用いて作製した。病理は浸潤度2の腺癌を示した。リンパ節には結腸外に広がった転移が見られた。患者の病歴には高脂血症、肉内疝および皮膚炎がある。家族歴には良性高血圧、アテローム硬化型冠動脈疾患、高脂血症、乳癌および前立腺癌がある。
CONNNOT01	pLINC	ライブラリは部分結腸切除術および永久的人工肛門形成術時(白人男性、71才)に得た腸間膜脂肪組織から単離したRNAを用いて作製した。家族歴にはアテローム硬化型冠動脈疾患、心筋梗塞および外閉性喘息がある。

10

20

30

40

【表19】

表6-2

ライブラリ	ベクター	ライブラリの説明
EOS1TUT01	pINCY	ライブラリは、IL-5で刺激した好球球から単離したRNAを用いて作製した。
GBLADIE01	PCDNA2.1	ライブラリは、ランダムプライムされたライブラリは55歳口人女性の腹腔鏡下胆嚢摘出術中に摘出された罹患胆嚢組織から単離したRNAを用いて作製した。病理は慢性胆嚢炎および胆石症(100個以上)を示した。患者は胆石症、腹部痛および嚢腫が見られた。患者の病歴には良性高血圧、モーターの神経腫、顔面多毛、正常な婦、および寛解期にあるニコチン中毒がある。以前の手術には腹式子宮全摘出術、両側性卵巣摘出術、およびアデノイド扁桃摘出術が含まれる。患者の服用薬剤はインテラルおよびプレマリンがある。家族歴には、母親に乳癌およびALS(筋萎縮性側索硬化症)、父親に慢性口血病およびARDS(急性呼吸窮迫症候群)があり、姉妹に乳癌、祖母にアテローム硬化型冠動脈疾患がある。
LATRFUT02	pINCY	ライブラリは43才白人男性の介輪形成術時に左心房から採取した培養したRNAを用いて作成した。病理は心房性粘液腫を示した。患者の病歴には肺不全、急性心筋梗塞、アテローム硬化型冠動脈疾患、高脂血症および喫煙がある。家族歴には良性高血圧症、急性心筋梗塞、アテローム硬化型冠動脈疾患およびII型糖尿病がある。

10

20

30

40

表7-1

プログラム	説明	引用文献	パラメータ関値
ABI FACTURA	核酸配列においてベクター配列を除去して、特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 AutoAssembler	核酸配列を構築するプログラム。	Applied Biosystems, Foster City, CA.	
BLAST	Basic Local Alignment Search Tool はアミノ酸および核酸配列の配列類似性検索に有用であり、BLAST はblastp, blastn, blastx, tblastnおよびtblastxの5つの機能がある。	Altschul, S.F. 他 (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. 他 (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: 確率値=1.0E-8 以下 完全長配列: 確率値1.0E-10 以下
FASTA	問合せ配列と同種の配列群との類似性を検索する Pearson およびLipman アルゴリズム。FASTAには最小5つの機能(fasta, ifasta, ifasta, ifastxおよびifastsearch)がある。	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: fasta E 値=1.0E-6 構築されたESTs: fasta 同一性=98%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: fasta スコア=100以上
BLIMPS	配列をBLOCKS、PRINTS、DOMO、PRODOMおよびPFAM データベースの配列と対応させて遺伝子ファミリー、配列相同性および構造的フィンガープリント領域を検索するBLOCKS IMPROVED サーチャャー。	Henikoff, S. および J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. および S. Henikoff (1996) Methods Enzymol. 266:88-105; および Attwood, T.K. 他 (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	確率値=1.0E-3以下
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; Sommarin, E.L.L. 他 (1988) Nucleic Acids Res. 26:320-322; Durbin, R. 他 (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM ヒット 確率値=1.0E-3 以下 シグナルペプチド ヒット: スコア=0 以上

10

20

30

40

【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
7 February 2002 (07.02.2002)

PCT

(10) International Publication Number
WO 02/10387 A2(51) International Patent Classification: C12N 15/12,
C07K 14/705, 16/28, C12Q 1/68, G01N 33/50, 33/577,
A61K 38/17, A01K 67/027

(21) International Application Number: PCT/US01/23433

(22) International Filing Date: 25 July 2001 (25.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/221,478 27 July 2000 (27.07.2000) US
60/223,268 3 August 2000 (03.08.2000) US
60/227,054 21 August 2000 (21.08.2000) US
60/231,121 8 September 2000 (08.09.2000) US
60/232,243 13 September 2000 (13.09.2000) US
60/232,691 15 September 2000 (15.09.2000) US
60/235,146 22 September 2000 (22.09.2000) 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): THORNTON,
Michael [US/US]; 9 Medway Road, Woodside, CA
94062 (US). PATTERSON, Chandra [US/US]; 490
Sherwood Way, #1, Menlo Park, CA 94025 (US). LAL,
Preeti [IN/US]; P.O. Box 5142, Santa Clara, CA 95056
(US). BURFORD, Nell [GB/US]; 105 Wildwood Circle,
Durham, CT 06422 (US). YUE, Henry [US/US]; 826 Lois
Avenue, Sunnyvale, CA 94087 (US). GANDHI, Ameena,
R. [US/US]; 837 Roble Avenue, #1, Redwood City, CA
94025 (US). ELLIOT, Vicki, S. [US/US]; 3770 Polton
Place Way, San Jose, CA 95121 (US). RAMKUMAR,
Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont,
CA 94555 (US). BAUGHN, Mariah, R. [US/US]; 14244
Santiago Road, San Leandro, CA 94577 (US). KALLICK,
Deborah, A. [US/US]; 900 Olive Street, Menlo Park, CA
94025 (US). WALIA, Narinder, K. [US/US]; 890 DavisStreet #205, San Leandro, CA 94577 (US). HAFALIA,
April, J., A. [US/US]; 2227 Calle de Primavera, Santa
Clara, CA 95054 (US). YAO, Monique, G. [US/US];
111 Frederick Court, Mountain View, CA 94043 (US).
LU, Yan [CN/US]; 3885 Corrina Way, Palo Alto, CA
94303 (US). TRIBOULEY, Catherine, M. [FR/US];
1121 Tennessee Street, #5, San Francisco, CA 94107
(US). POLICKY, Jennifer, L. [US/US]; 1511 Jarvis
Court, San Jose, CA 95118 (US). KEARNEY, Liam
[IE/US]; 50 Woodside Avenue, San Francisco, CA 94127
(US). GRAUL, Richard, C. [—/US]; 682-29th Avenue,
San Francisco, CA 94121 (US). WARREN, Bridget, A.
[US/US]; 10130 Parwood Drive #2, Cupertino, CA 95014
(US). LEE, Ernestine, A. [US/US]; 624 Kains Street,
Albany, CA 94706 (US). DING, Li [US/US]; 3353 Alma
Street, #146, Palo Alto, CA 94306 (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,
TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BI, CF,
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG).

Published:

— without international search report and to be republished
upon receipt of that reportFor two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(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, agonist, and antagonist. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of GCREC.



WO 02/10387 A2

WO 02/10387

PCT/US01/23433

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 02/10387

PCT/US01/23433

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 transmembrane domains are less common. Some receptors have more than one site at which variance
20 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 metabotropic glutamate receptor subfamilies. Members of these GPCR subfamilies share similar
25 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 stacks found in photoreceptor (rod) cells. Each rhodopsin molecule responds to a photon of light by
30 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 02/10387

PCT/US01/23433

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
5 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
15 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
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 02/10387

PCT/US01/23433

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

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

GPCR mutations, which may cause loss of function or constitutive activation, have been associated with numerous human diseases (Coughlin, *supra*). For instance, retinitis pigmentosa may arise from mutations in the rhodopsin gene. Furthermore, somatic activating mutations in the thyrotropin receptor have been reported to cause hyperfunctioning thyroid adenomas, suggesting that certain GPCRs susceptible to constitutive activation may behave as protooncogenes (Parma, J. et al. (1993) Nature 365:649-651). GPCR receptors for the following ligands also contain mutations associated with human disease: luteinizing hormone (precocious puberty); vasopressin V₂ (X-linked nephrogenic diabetes); glucagon (diabetes and hypertension); calcium (hyperparathyroidism, hypocalcemia, hypercalcemia); parathyroid hormone (short limbed dwarfism); β_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 02/10387

PCT/US01/23433

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 02/10387

PCT/US01/23433

collectively as "GCREC" and individually as "GCREC-1," "GCREC-2," "GCREC-3," "GCREC-4," "GCREC-5," "GCREC-6," "GCREC-7," "GCREC-8," "GCREC-9," "GCREC-10," "GCREC-11," "GCREC-12," "GCREC-13," "GCREC-14," "GCREC-15," "GCREC-16," "GCREC-17," "GCREC-18," and "GCREC-19." 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:1-19, 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:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-19.

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:1-19, 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:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-19. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:20-38.

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:1-19, 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:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. 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:1-19, b) a polypeptide comprising a naturally occurring amino acid sequence at least

WO 02/10387

PCT/US01/23433

90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. The method comprises a) 5 culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid 10 sequence selected from the group consisting of SEQ ID NO:1-19, 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:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group 15 consisting of SEQ ID NO:1-19.

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:20-38, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at 20 least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, 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 25 consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, 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:20-38, 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 30 method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if 35 present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous

WO 02/10387

PCT/US01/23433

nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, 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:20-38, 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:1-19, 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:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. 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:1-19, 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:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. 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,

WO 02/10387

PCT/US01/23433

comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, b) a polypeptide
5 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:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. The method comprises a) exposing a sample comprising the
10 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.

15 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:1-19, 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:1-19, c) a biologically active fragment of a polypeptide
20 having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. 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.

25 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:1-19, 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:1-19, c) a biologically active fragment of a polypeptide
30 having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. 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

WO 02/10387

PCT/US01/23433

polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, 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:20-38, 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:20-38, 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:20-38, 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:20-38, 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.

WO 02/10387

PCT/US01/23433

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

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

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.

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.

WO 02/10387

PCT/US01/23433

DEFINITIONS

"GCREC" refers to the amino acid sequences of substantially purified GCREC obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

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

10 An "allelic variant" is an alternative form of the gene encoding GCREC. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times
15 in a given sequence.

"Altered" nucleic acid sequences encoding GCREC include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as GCREC or a polypeptide with at least one functional characteristic of GCREC. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide
20 probe of the polynucleotide encoding GCREC, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding GCREC. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent GCREC. Deliberate amino acid substitutions may be made on the basis of similarity in
25 polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of GCREC is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine.
30 Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

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

WO 02/10387

PCT/US01/23433

sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence.

Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

5 The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of GCREC. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of GCREC either by directly interacting with GCREC or by acting on components of the biological pathway in which GCREC participates.

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

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

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

WO 02/10387

PCT/US01/23433

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic GCREC, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding GCREC or fragments of GCREC may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

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

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

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

WO 02/10387

PCT/US01/23433

	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
5	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
10	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.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of 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

WO 02/10387

PCT/US01/23433

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.

10 A fragment of SEQ ID NO:20-38 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:20-38, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:20-38 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:20-38 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:20-38 and the region of SEQ ID NO:20-38 to which the fragment corresponds are routinely
15 determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-19 is encoded by a fragment of SEQ ID NO:20-38. A fragment of SEQ ID NO:1-19 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-19. For example, a fragment of SEQ ID NO:1-19 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-19. The precise length of
20 a fragment of SEQ ID NO:1-19 and the region of SEQ ID NO:1-19 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

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

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

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

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e
35 sequence alignment program. This program is part of the LASERGENE software package, a suite of

WO 02/10387

PCT/US01/23433

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

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

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

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

35 Nucleic acid sequences that do not show a high degree of identity may nevertheless encode

WO 02/10387

PCT/US01/23433

similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

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

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

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

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

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

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

WO 02/10387

PCT/US01/23433

chromosome replication, segregation and maintenance.

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

5 "Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the
10 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.
15 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
20 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.

25 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
30 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
35 similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

WO 02/10387

PCT/US01/23433

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C₀t or R₀t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

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

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

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of GCREC which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of GCREC which is useful in any of the antibody production methods disclosed herein or known in the art.

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

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

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

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

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

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of

WO 02/10387

PCT/US01/23433

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.

5 "Post-translational modification" of an GCREC may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of GCREC.

"Probe" refers to nucleic acid sequences encoding GCREC, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are 10 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 15 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 20 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 25 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).

30 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 35 PrimOU primer selection program (available to the public from the Genome Center at University of

WO 02/10387

PCT/US01/23433

Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence 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

WO 02/10387

PCT/US01/23433

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

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

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

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

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

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

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

WO 02/10387

PCT/US01/23433

replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

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

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

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of

WO 02/10387

PCT/US01/23433

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
5 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 (GCREC), the polynucleotides encoding GCREC, and the use of these compositions for the diagnosis,
10 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
15 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
20 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
25 homolog. Column 5 shows the annotation of the GenBank homolog.

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
30 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.

35 Together, Tables 2 and 3 summarize the properties of each polypeptide of the invention, and

WO 02/10387

PCT/US01/23433

these properties establish that the claimed polypeptides are G-protein coupled receptors. For example, SEQ ID NO:1 is 40% identical to Meleagris gallopavo G protein-coupled P2Y nucleotide receptor (GenBank ID g2707256) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 4.0e-62, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a rhodopsin family 7 transmembrane receptor 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 and BLAST analyses provide further corroborative evidence that SEQ ID NO:1 is G-protein coupled receptor. SEQ ID NO:2 was analyzed and annotated in a similar manner. These analyses indicate that SEQ ID NO:2 is a pheromone receptor (Dulac, C. and R. Axel (1995) Cell 83:195-206).

As a further example, SEQ ID NO:6 is 29% identical to human C-C chemokine receptor type 1 (GenBank ID g179985) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.6e-15, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also contains a 7 transmembrane receptor (rhodopsin family) domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:6 is a chemokine receptor.

As a further example, SEQ ID NO:9 is 95% identical to rat calcium-independent alpha-latrotoxin receptor (GenBank ID g3882981) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:9 also contains a 7-transmembrane receptor (secretin family) domain and a latrophilin/CL-1-like GPS domain, as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:9 is a latrophilin-related G-protein coupled receptor.

As a further example, SEQ ID NO:12 is 84% identical to Mus musculus G-protein coupled receptor GPR73 (GenBank ID g7248884) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 6.7e-166, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:12 also contains a 7 transmembrane receptor (rhodopsin family) domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS analysis reveals the presence of a rhodopsin-like

WO 02/10387

PCT/US01/23433

GPCR superfamily signature (See Table 3). Additional data from MOTIFS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:12 is a G-protein coupled receptor.

As a further example, SEQ ID NO:15 is 80% identical to rat serotonin receptor (GenBank ID g310075) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $2.5e-152$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:15 also contains a rhodopsin family receptor domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, analyses provide further corroborative evidence that SEQ ID NO:15 is a G-protein coupled receptor.

As a further example, SEQ ID NO:16 is 71% identical to mouse olfactory receptor E3 (GenBank ID g3983382) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $1.9e-88$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:16 also contains a rhodopsin family 7-transmembrane receptor domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:16 is an olfactory G-protein coupled receptor.

As a further example, SEQ ID NO:17 is 83% identical to mouse olfactory G-protein coupled receptor G3 (GenBank ID g3983398) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $5.0e-99$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:17 also contains a rhodopsin family 7-transmembrane receptor domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:17 is an olfactory G-protein coupled receptor. SEQ ID NO:2-5, SEQ ID NO:7-8, SEQ ID NO:10-11, SEQ ID NO:13-14, and SEQ ID NO:18-19 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-19 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

WO 02/10387

PCT/US01/23433

of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:20-38 or that distinguish between SEQ ID NO:20-38 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, 7075196H1 is the identification number of an Incyte cDNA sequence, and BRAUTDR04 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., 71906055V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g900324) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the 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, FL_XXXXXX_N₁_N₂_YYYY_N₃_N₄ represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYY is the number of the prediction generated by the algorithm, and N_{1,2,3,4}, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FL_XXXXXX_gAAAAA_gBBBBB_1_N is the identification number of a "stretched" sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

WO 02/10387

PCT/US01/23433

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

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

The invention also encompasses GCREC variants. A preferred GCREC variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the GCREC amino acid sequence, and which contains at least one functional or structural characteristic of GCREC.

The invention also encompasses polynucleotides which encode GCREC. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:20-38, which encodes GCREC. The polynucleotide sequences of SEQ ID NO:20-38, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding GCREC. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at

WO 02/10387

PCT/US01/23433

least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding GCREC. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:20-38 which has at least about 70%, or alternatively at least about 85%, or even at least about 5 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:20-38. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of GCREC.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding GCREC, some bearing minimal 10 similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring GCREC, and all such variations are to be considered 15 as being specifically disclosed.

Although nucleotide sequences which encode GCREC and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring GCREC under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding GCREC or its derivatives possessing a substantially different codon usage, e.g., inclusion of non- 20 naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding GCREC and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a 25 greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode GCREC and GCREC derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to 30 introduce mutations into a sequence encoding GCREC or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:20-38 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.* 35 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in

WO 02/10387

PCT/US01/23433

"Definitions."

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

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

WO 02/10387

PCT/US01/23433

about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

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

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode GCREC may be cloned in recombinant DNA molecules that direct expression of GCREC, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express GCREC.

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

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of GCREC, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired

WO 02/10387

PCT/US01/23433

properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding GCREC may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser.* 7:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser.* 7:225-232.) Alternatively, GCREC itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) *Proteins, Structures and Molecular Properties*, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of GCREC, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

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

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

WO 02/10387

PCT/US01/23433

frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

5 Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding GCREC and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

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

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

WO 02/10387

PCT/US01/23433

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. 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.

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) *Bio/Technology* 12:181-184.)

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

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

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.)

WO 02/10387

PCT/US01/23433

For long term production of recombinant proteins in mammalian systems, stable expression of GCREC in cell lines is preferred. For example, sequences encoding GCREC can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector.

5 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.

10 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.

20 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.)

25 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

30 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

35 amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or

WO 02/10387

PCT/US01/23433

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

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

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding GCREC include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding GCREC, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding GCREC may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode GCREC may be designed to contain signal sequences which direct secretion of GCREC through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for

WO 02/10387

PCT/US01/23433

post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WB38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

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

WO 02/10387

PCT/US01/23433

compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express GCREC, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing GCREC or cell membrane fractions which contain GCREC are then
5 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
10 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
20 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
25 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
30 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
35 place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-

WO 02/10387

PCT/US01/23433

specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding 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 brain tissue, fetal brain tissue, colon polyps, diseased colon tissue, colon tumor tissue, diseased gallbladder tissue, heart tissue, diseased breast tissue, interleukin-5 stimulated eosinophils, tumor tissue, and reproductive tissues. 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

WO 02/10387

PCT/US01/23433

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, colon, gall bladder, ganglia, 5 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, 10 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 familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, 15 tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, 20 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 cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, 25 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, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular 30 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 as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma,

WO 02/10387

PCT/US01/23433

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 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, hyper eosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a metabolic disorder such as diabetes, obesity, and osteoporosis; and an infection by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepatitis virus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and togavirus.

In another embodiment, a vector capable of expressing GCREC or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC including, but not limited to, those described above.

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

WO 02/10387

PCT/US01/23433

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.

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

10 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.

15 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.

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

25 An antagonist of 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

WO 02/10387

PCT/US01/23433

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

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to GCREC have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of GCREC amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to GCREC may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce GCREC-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

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

Antibody fragments which contain specific binding sites for GCREC may also be generated. For example, such fragments include, but are not limited to, F(ab)₂ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab)₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either

WO 02/10387

PCT/US01/23433

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

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for GCREC. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of GCREC-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple GCREC epitopes, represents the average affinity, or avidity, of the antibodies for GCREC. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular GCREC epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the GCREC-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a 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.)

WO 02/10387

PCT/US01/23433

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 5 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 10 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 15 (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* 20 *Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassamias, 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 25 (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 30 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 35 use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii)

WO 02/10387

PCT/US01/23433

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) *Ann. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J.-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

- 5 Expression vectors that may be effective for the expression of GCRC include, but are not limited to, the pCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). GCRC may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV),
- 10 Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-5551; Gossen, M. et al. (1995) *Science* 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) *Curr. Opin. Biotechnol.* 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND;
- 15 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 TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver

20 polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

- 25 In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to 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 element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences
- 30 required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al.
- 35 (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and

WO 02/10387

PCT/US01/23433

A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by
5 reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding GCREC to cells which have one or more genetic abnormalities with respect to the expression of GCREC. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to
15 be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Cséte, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) *Annu. Rev. Nutr.* 19:511-544 and Verma, I.M. and N. Somia (1997) *Nature* 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding GCREC to target cells which have one or more genetic abnormalities with respect to the expression of GCREC. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing GCREC to cells of the central nervous system, for which HSV has
25 a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is
30 hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned
35

WO 02/10387

PCT/US01/23433

herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

5 In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding GCREC to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid 10 proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for GCREC into the alphavirus genome in place of the capsid-coding region results in the production of a large number of GCREC-coding RNAs and the synthesis of high levels of GCREC in vector transduced cells. While 15 alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of GCREC into a variety of cell types. The specific transduction of a subset of 20 cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions 25 -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.J. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-30 177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme 35 molecule to complementary target RNA, followed by endonucleolytic cleavage. For example,

WO 02/10387

PCT/US01/23433

engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding GCREC.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, 5 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.

10 Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding GCREC. Such DNA sequences may be incorporated into a wide variety of 15 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' 20 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 25 endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding GCREC. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming 30 oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased GCREC expression or activity, a compound which specifically inhibits expression of the 35 polynucleotide encoding GCREC may be therapeutically useful, and in the treatment of disorders

WO 02/10387

PCT/US01/23433

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

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

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

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

WO 02/10387

PCT/US01/23433

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

The compositions utilized in this invention may be administered by any number of routes 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. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without 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.

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 transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) *Science* 285:1569-1572).

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

WO 02/10387

PCT/US01/23433

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

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

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

WO 02/10387

PCT/US01/23433

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
5 extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to GCREC under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of GCREC expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for
10 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
15 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
20 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%
25 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:20-38 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
30 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.

35 Polynucleotide sequences encoding GCREC may be used for the diagnosis of disorders

WO 02/10387

PCT/US01/23433

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

5 adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, 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

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

15 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

20 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, cataplexy, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic

25 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,

30 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

WO 02/10387

PCT/US01/23433

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

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

10 obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha₁-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein

15 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,

20 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, hypercosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis,

25 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

30 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 GCREC may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like

WO 02/10387

PCT/US01/23433

assays; and in microarrays utilizing fluids or tissues from patients to detect altered GCREC expression. Such qualitative or quantitative methods are well known in the art.

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

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

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

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

Additional diagnostic uses for oligonucleotides designed from the sequences encoding GCREC may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide

WO 02/10387

PCT/US01/23433

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.

5 In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding GCREC may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, 10 oligonucleotide primers derived from the polynucleotide sequences encoding GCREC are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the 15 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 20 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 GCREC include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from 25 standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; 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.

30 In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene 35 function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor

WO 02/10387

PCT/US01/23433

progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

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

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

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

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) *Mol. Carcinog.* 24:153-159; Steiner, S. and N.L. Anderson (2000) *Toxicol. Lett.* 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is

WO 02/10387

PCT/US01/23433

not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

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

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by 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

WO 02/10387

PCT/US01/23433

polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for GCREC to quantify the levels of GCREC expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoz, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

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

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

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) *Proc. Natl. Acad. Sci.*

WO 02/10387

PCT/US01/23433

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. 5 (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding GCREC may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among 10 members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. 15 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- 20 7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding GCREC on a 25 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, 30 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 35 investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of

WO 02/10387

PCT/US01/23433

the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, GCREC, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between GCREC and the agent being tested may be measured.

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 preferred specific 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, in particular U.S. Ser. No. 60/221,478, U.S. Ser. No. 60/223,268, U.S. Ser. No. 60/231,121, U.S. Ser. No. 60/232,691, U.S. Ser. No. 60/235,146, U.S. Ser. No. 60/227,054, and U.S. Ser. No. 60/232,243, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

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

WO 02/10387

PCT/US01/23433

(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.

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

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPIT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column 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., PBLUESCRIPIT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

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

WO 02/10387

PCT/US01/23433

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

III. Sequencing and Analysis

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

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

WO 02/10387

PCT/US01/23433

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

5 GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the

10 CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of 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,

15 the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

20 The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:20-38. 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

25 Putative G-protein coupled receptors were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhitg). 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

30 form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode G-protein coupled receptors, the encoded polypeptides were analyzed by querying against PFAM models for G-protein coupled receptors. Potential G-protein

35 coupled receptors were also identified by homology to Incyte cDNA sequences that had been

WO 02/10387

PCT/US01/23433

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

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

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

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public

WO 02/10387

PCT/US01/23433

databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of GREC Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:20-38 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:20-38 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

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

WO 02/10387

PCT/US01/23433

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:

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

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

Alternatively, polynucleotide sequences encoding GCRC are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; 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

WO 02/10387

PCT/US01/23433

disease-specific expression of cDNA encoding GCRC. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of GCRC Encoding Polynucleotides

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

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg²⁺, (NH₄)₂SO₄, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1% agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviII cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%)

WO 02/10387

PCT/US01/23433

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

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

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

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:20-38 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 µCi of [γ -³²P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10⁷ counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and

WO 02/10387

PCT/US01/23433

compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schna (1999), *supra*). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schna, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31.)

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

25 Tissue or Cell Sample Preparation

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

WO 02/10387

PCT/US01/23433

using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and
5 resuspended in 14 μ l 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are
10 amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water
15 washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US
20 Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average concentration of 100 ng/ μ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water.
25 Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μ l of sample mixture consisting of 0.2 μ g each of Cy3 and
30 Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is

WO 02/10387

PCT/US01/23433

incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

5 Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each

WO 02/10387

PCT/US01/23433

spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

5 Sequences complementary to the GCREC-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring GCREC. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of GCREC.

10 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

15 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

20 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

25 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.

30 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-

35 kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on

WO 02/10387

PCT/US01/23433

immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from GCREC at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified GCREC obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, 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

WO 02/10387

PCT/US01/23433

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

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

Alternatively, the GCREC amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is 10 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- 15 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 20 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 25 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 30 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

GCREC, or biologically active fragments thereof, are labeled with ¹²⁵I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules 35 previously arrayed in the wells of a multi-well plate are incubated with the labeled GCREC, washed,

WO 02/10387

PCT/US01/23433

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 molecules.

Alternatively, molecules interacting with GCREC are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

GCREC may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Demonstration of GCREC Activity

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

In the alternative, an assay for GCREC activity is based on a prototypical assay for ligand/receptor-mediated modulation of cell proliferation. This assay measures the rate of DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding GCREC is added to quiescent 3T3 cultured cells using transfection methods well known in the art. The transiently transfected cells are then incubated in the presence of [³H]thymidine, a radioactive DNA precursor molecule. Varying amounts of GCREC ligand are then added to the cultured cells. Incorporation of [³H]thymidine into acid-precipitable DNA is measured over an appropriate time interval using a radioisotope counter, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold GCREC ligand concentration range is indicative of receptor activity. One unit of activity per milliliter is defined as the concentration of GCREC producing a 50% response level, where 100% represents maximal incorporation of [³H]thymidine into acid-precipitable DNA (McKay, I. and I. Leigh, eds. (1993) *Growth Factors: A Practical Approach*, Oxford University Press, New York NY, p. 73.)

In a further alternative, the assay for GCREC activity is based upon the ability of GPCR family proteins to modulate G protein-activated second messenger signal transduction pathways (e.g., cAMP; Gaudin, P. et al. (1998) *J. Biol. Chem.* 273:4990-4996). A plasmid encoding full length

WO 02/10387

PCT/US01/23433

GCREC is transfected into a mammalian cell line (e.g., Chinese hamster ovary (CHO) or human embryonic kidney (HEK-293) cell lines) using methods well-known in the art. Transfected cells are grown in 12-well trays in culture medium for 48 hours, then the culture medium is discarded, and the attached cells are gently washed with PBS. The cells are then incubated in culture medium with or without ligand for 30 minutes, then the medium is removed and cells lysed by treatment with 1 M perchloric acid. The cAMP levels in the lysate are measured by radioimmunoassay using methods well-known in the art. Changes in the levels of cAMP in the lysate from cells exposed to ligand compared to those without ligand are proportional to the amount of GCREC present in the transfected cells.

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

XVIII. Identification of GCREC Ligands

GCREC is expressed in a eukaryotic cell line such as CHO (Chinese Hamster Ovary) or HEK (Human Embryonic Kidney) 293 which have a good history of GPCR expression and which contain a wide range of G-proteins allowing for functional coupling of the expressed GCREC to downstream effectors. The transformed cells are assayed for activation of the expressed receptors in the presence of candidate ligands. Activity is measured by changes in intracellular second messengers, such as cyclic AMP or Ca^{2+} . These may be measured directly using standard methods well known in the art, or by the use of reporter gene assays in which a luminescent protein (e.g. firefly luciferase or green fluorescent protein) is under the transcriptional control of a promoter responsive to the stimulation of protein kinase C by the activated receptor (Milligan, G. et al. (1996) Trends Pharmacol. Sci. 17:235-237). Assay technologies are available for both of these second messenger systems to allow high throughput readout in multi-well plate format, such as the adenylyl cyclase activation FlashPlate Assay (NEN Life Sciences Products), or fluorescent Ca^{2+} indicators such as Fluo-4 AM (Molecular Probes) in combination with the FLIPR fluorimetric plate reading system (Molecular Devices). In cases where the physiologically relevant second messenger pathway is not known, GCREC may be coexpressed with the G-proteins $G_{\alpha_{15/16}}$ 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 the signal transduction of the GCREC through a pathway involving phospholipase C and Ca^{2+}

WO 02/10387

PCT/US01/23433

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 Ga protein for the corresponding components of the endogenous yeast pheromone receptor pathway. Downstream signaling pathways are also modified so that the normal yeast response to the signal is converted to positive growth on selective media or to reporter gene expression (Broach, J.R. and J. Thorner (1996) Nature 384 (supp.):14-16). The receptors are screened against putative ligands including known GPCR ligands and other naturally occurring bioactive molecules. Biological extracts from tissues, biological fluids and cell supernatants are also screened.

10

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

15

WO 02/10387

PCT/US01/23433

Table 1

Project ID	Incyte Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Incyte Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
747806	1	747806CD1	20	747806CB1
747807	2	747807CD1	21	747807CB1
747808	3	747808CD1	22	747808CB1
747809	4	747809CD1	23	747809CB1
7341260	5	7341260CD1	24	7341260CB1
7473911	6	7473911CD1	25	7473911CB1
7474767	7	7474767CD1	26	7474767CB1
7475815	8	7475815CD1	27	7475815CB1
60263275	9	60263275CD1	28	60263275CB1
60203310	10	60203310CD1	29	60203310CB1
7477349	11	7477349CD1	30	7477349CB1
55002225	12	55002225CD1	31	55002225CB1
7475686	13	7475686CD1	32	7475686CB1
7482007	14	7482007CD1	33	7482007CB1
5762042	15	5762042CD1	34	5762042CB1
7481940	16	7481940CD1	35	7481940CB1
55036418	17	55036418CD1	36	55036418CB1
7481701	18	7481701CD1	37	7481701CB1
7481774	19	7481774CD1	38	7481774CB1

WO 02/10387

PCT/US01/23433

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Genbank ID NO:	Probability score	Genbank Homolog
1	7474806CD1	g2707256	4.00E-62	[f1][Mus musculus] gallopavo G protein coupled P2Y nucleotide receptor
2	7474840CD1	g11967419	4.00E-65	[f1][Mus musculus] vomeronasal (pheromone) receptor V1RC3
3	7475092CD1	g2992628	2.00E-86	[f1][Homo sapiens] putative seven pass transmembrane protein
4	7341260CD1	g3805932	1.00E-10	[f1][Homo sapiens] putative G-Protein coupled receptor, EDG6 Graler, M.H. et al. (1998) Genomics 53:164-169
5	7473941CD1	g1055254	2.00E-18	[f1][Rattus norvegicus] pheromone receptor VM6 Dulac, C. and Axel, R. (1995) Cell 83: 293-206
6	7474767CD1	g179985	2.00E-14	[f1][Homo sapiens] C-C chemokine receptor type 1 Meste, K. et al. (1993) Cell 72:415-425
7	7475815CD1	g1055254	1.00E-56	[f1][Rattus norvegicus] pheromone receptor VM6
8	60263273CD1	g13183149	0	[f1][Homo sapiens] (AF239764) EGF-like module-containing mucin-like receptor EMR3
9	60203310CD1	g3882981	0	[f1][Rattus norvegicus] calcium-independent alpha-latrotoxin receptor
10	7477349CD1	g6979162	3.00E-11	[f1][Rattus norvegicus] macrophage scavenger receptor type 1 receptor
11	55002225CD1	g14164383	0	[f1][Homo sapiens] (AB060151) G protein-coupled receptor
12	7475686CD1	g7248884	1.00E-130	[f1][Mus musculus] G-protein coupled receptor GPR73 Parker, R. et al. (2000) Biochim. Biophys. Acta 1491:369-375
13	7482007CD1	g5525078	1.00E-104	[f1][Rattus norvegicus] seven transmembrane receptor Kobayashi, T. et al. (1999) J. Biol. Chem. 274:19957-19964
14	6769042CD1	g4164061	1.70E-59	[Bos taurus] lactophilin 3 splice variant abbg FEBS Lett. (1999) 443:348-352

WO 02/10387

PCT/US01/23433

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
15	747605CD1	g310075	1.00E-165	[Fl][Rattus norvegicus] serotonin receptor Prilander, M.G. et al. (1993) Proc. Natl. Acad. Sci., U.S.A. 90:3452-3456
16	7480410CD1	g3983382	1.90E-88	[Mus musculus] olfactory receptor E3
17	55036418CD1	g3983398	5.00E-99	[Mus musculus] olfactory receptor G3 Krautwurst, D. et al. (1998) Cell 95:917-926
18	7481701CD1	g12007416	2.00E-67	[Fl][Mus musculus] M51 olfactory receptor
19	7481774CD1	g5901478	3.10E-106	[Marmota marmota] olfactory receptor

WO 02/10387

PCT/US01/23433

Table 3

SEQ ID NO.	Incyte Peptide	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Reference Databases
1	7474806CD1	339	S113 S15 S189 S260 S336 S46 S84 T163 T204 T255 T36 T80	M35 N69	Transmembrane domains: Y89-T109, F130-F149 family: L68-Y324 7 transmembrane receptor (rhodopsin family) G-protein coupled receptor BL00237: M116-P155, F224-Y235, T255-Y281, S316-I332 Rhodopsin-like GPCR superfamily PR00237: Y53-W77, A86-I107, F130-V152, H166-A187, G216-Y239, S260-F284, X306-I332 G-protein coupled receptor: DR00013 S113 S15 S189 S260 S336 S46 S84 T163 T204 T255 T36 T80 F15482 29-322: F45-S327 I55450 20-317: F45-R328 F48042 45-340: Y53-S327 G-protein coupled receptor: PD00009: Y89-F188	HMMER-FPAM HMMER-FPAM ELIMPS-BLOCKS ELIMPS-PRINTS ELAST-DOMO ELAST-PRODOM ELAST-PRODOM
2	7474840CD1	335	S237 S28 S118 T22 T86 S237 S286 S295 S412 S79 T158	M168 M184	Pharomane receptor: PD009900: I59-M329	ELAST-PRODOM
3	7475092CD1	428	S237 S286 S295 S412 S79 T158	M180 M207 M34	Transmembrane domain: F130-F149, M240-V256, Y299-I319 Putative seven-pass transmembrane protein: PD138976: R3-I360	HMMER ELAST-PRODOM
4	7341260CD1	330	S123 S185 S323 S5 S78 T219	M154 M4 N76 M93	Transmembrane domain: L55-L74 family: A31-A250 7 transmembrane receptor (rhodopsin family) Rhodopsin-like GPCR superfamily signature PR00237B: A50-L71 PR00237C: P92-V114 PR00237D: S123-L144 PR00237F: Q222-L246 PR00237G: L262-T288	HMMER HMMER-FPAM ELIMPS-PRINTS
5	7473911CD1	676	S105 S213 S215 S250 S330 S353 S492 S569 S591 S606 S646 S647 S79 T217 T296 T301 T394 T620	M184	Transmembrane domain: I118-A136 PHEROMONE RECEPTOR PD009900: S87-V167, K42-Y90	HMMER ELAST-PRODOM

WO 02/10387

PCT/US01/23433

Table 3 (cont.)

SEQ ID NO	Incyte Polypeptide	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	7474767CD1	372	S214 S289 S351 T145 T223 T239	N11 N318 N353	Signal Peptide: N1-R47 Transmembrane Domains: Y32-S52, I74-F100, I184-Y208 Transmembrane receptor (rhodopsin family): A43-Y283 G-protein coupled receptors proteins R20237:R0237, R221-R247, R277-R293 Rhodopsin-like GPCR superfamily signatures R09097:V38-S52, S62-64, E108-E130, RL44-ML65, I184-Y207, Y226-Y239, S267-R293 G-protein coupled receptors signature: R109-F155 S-PROTEIN COUPLED RECEPTORS DM00013 P32246 28-316:A22-R293 P51677 28-316:F27-R293 P25930 33-319:F27-L500 P34961 19-336:V28-L206, S222-R293	SSSCAN HMMER HMMER_PPAM BLIHP5_BLOCKS BLIHP5_PRINTS
7	7475815CD1	271	S115 S142 S188 S203 S79 T269		Transmembrane Domain: I216-A234 PHARMOPHORE RECEPTOR F009900: K42-S170, I112-K271	HMMER BLAST_PRODOM BLAST_DOMO

WO 02/10387

PCT/US01/23433

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
8	60263275CD1	611	S237 S308 S494 S578 S580 S588 S598 S63 S81 T21 T211 T238 T307 T406 T75 Y585	N104 N148 N161 N209 N238 N286 N293 N345 N409 N414	Signal Peptide: M1-Q22 Transmembrane Domains: V321-L339, W424-L451, M464-L488, Y340-L560 7 transmembrane receptor (Secretin Family): D312-Y564 Latrophilin/CU-1-like GFS domain: K259-Q309 EGF-like domain: C30-C76 G-protein coupled receptor BL00649: C378-L403, G430-R454, M465-S494 Secretin-like GPCR superfamily signature PR00249: V317-K341, E380-L403, K423-L448 M465-K490, A539-L560 Type I EGF signature PR00009: K25-F40, E48-Y59 G-protein coupled receptors family 2 signatures: V531-E574 RECEPTOR G PROTEIN COUPLED PD000752: L315-F572 RECEPTOR G PROTEIN COUPLED EGF LIKE PD05428: M176-M304 HORMONE; EMRI; LEUCOCYTE; ANTIGEN; DR05221 A57172 465-886: S183-K596 I37225 347-738: K259-G552 P48960 347-738: K259-G552 G-PROTEIN COUPLED RECEPTORS FAMILY 2 DR00378 I49149 56-425: V314-R573, G9-V56 Aspartic acid and asparagine hydroxylation site: C44-C55	HOMER HOMER HOMER_PPAM HOMER_PPAM BLIMPS_BLOCKS BLIMPS_PRINTS BLIMPS_PRINTS BLIMPS_PRINTS PROFILESSCAN BLAST_PRODOM BLAST_PRODOM BLAST_DOMO BLAST_DOMO MOTIFS

WO 02/10387

PCT/US01/23433

Table 3 (cont.)

SEQ ID NO:	Incyte Pelypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	60203310CDD1	1489	S166 S186 S194 S225 S235 S251 S1295 N837 N843 N931 S417 S423 T1198 N932 N1098 S460 S591 T1453 N1179 N1196 S630 S761 T1008 N1251 S791 S824 S1144 S88 S907 S1154 S174 T193 S1395 T234 T303 T1346 T443 T559 S1458 T658 T672 T731 T803 T842 Y121	N164 N549 N672 N843 N931 N1098 N1179 N1196 N1251	Signal peptide: M1-S21 7 Transmembrane receptor (secretin family): D874-V1130 Lactophilin/CLL-like GFS domain: F114-V856 G-protein coupled receptor BL00649: C940-L965, G987-V1011, M1022-A1051, S1113-L1138, C498-T525, G884-L929 Secretin-like GPCR superfamily PR00249: V948-L965, G987-V1011, M1022-A1051, S1113-L1138, C498-T525, G884-L929 TMR1 hormone leucocyte antigen: DM05231 P48960 347-738: L769-S1160 A57172 465-886: N730-N1162 DM00378 149149156-425: G836-R1131 G-protein coupled receptors Family 2: DM00378 149149156-425: G836-R1131 Lactophilin-related receptor: L1207-L1469 PD024331: K386-B828 PD021747: L1207-L1469 G-protein coupled receptor response protein: PD068897, V127-V385 EGFR-like G-protein coupled receptor: PD005428: R629-V856 Transmembrane domains: F166-L188, V215-V233 7 Transmembrane receptor (rhodopsin family): A251-Y421 G-protein coupled receptor motif: A951-L267 G-protein coupled receptors signature: I243-A293 G-protein coupled receptor BL00237: R331-G709 G-protein coupled receptor BL00238: H112-R409 EGFR-like GPCR superfamily: T165-A189, S199-A220, E245-L267, R281-W302, K331-I344, S362-Y386, L403-R429 G-protein coupled receptor: DM00013 I49339 28-316: A162-V435 P51675 128-317: A162-V435	HMMER HMMER HMMER-PFAM HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-DMO BLAST-DMO BLAST-DMO BLAST-PRODOM BLAST-PRODOM BLAST-PRODOM HMMER HMMER-PFAM MOTIFS ProfileScan BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-DMO BLAST-DMO
10	7477349CDD1	469	S122 S199 S225 S312 S447 S85 T110 T195 T207 T282 T315 T319 T427 T431 T465	N139	Transmembrane domains: F166-L188, V215-V233 7 Transmembrane receptor (rhodopsin family): A251-Y421 G-protein coupled receptor motif: A951-L267 G-protein coupled receptors signature: I243-A293 G-protein coupled receptor BL00237: R331-G709 G-protein coupled receptor BL00238: H112-R409 EGFR-like GPCR superfamily: T165-A189, S199-A220, E245-L267, R281-W302, K331-I344, S362-Y386, L403-R429 G-protein coupled receptor: DM00013 I49339 28-316: A162-V435 P51675 128-317: A162-V435	HMMER HMMER-PFAM MOTIFS ProfileScan BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-DMO BLAST-DMO

WO 02/10387

PCT/US01/23433

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 Domains
11	5500225CD1	335	S230 S269 S601 S607 T243 T313 T117 T110 T150 T189 T320 T330 T66	N10 N17	G-PROTEIN COUPLED RECEPTORS EM000131:P32745138-429 W32-S310 RECEPTOR COUPLED G-PROTEIN TRANSMEMBRANE GLYCOPROTEIN PHOSPHORYLATION LIPOPROTEIN PALMITATE PROTEIN FAMILY PD00009:Rgi-Y171 G-protein coupled receptor BL00237:IS298- Q314_W99-F138_E206-Y217_Q244-I270 Rhodopsin-like GPCR superfamily signature PR00237:136-160_F68-L69_T113-L135_X149- Y170_T188-L221_W48-L52Z737_Y236-Q214 Transmembrane domain:R34- E57_V194-T218_W352-L272 7 transmembrane receptor (rhodopsin family) 7tm_1:G51-Y306 G-Protein Receptor A119-L135 G-protein coupled receptors signature S-protein_receptor.prf.M10-L157	BLAST_DOMO BLAST_PRODUM BLIMPS_BLOCKS BLIMPS_PRINTS HMMER HMMER_PPFAM MOTIFS PROFILESSCAN
12	7475686CD1	630	S230 S269 S601 S607 T243 T313 T81 T366 T620 T81	N332 N584	G-PROTEIN COUPLED RECEPTORS EM000131:P32745138-429 W32-S310 RECEPTOR COUPLED G-PROTEIN TRANSMEMBRANE GLYCOPROTEIN PHOSPHORYLATION LIPOPROTEIN PALMITATE PROTEIN FAMILY PD00009:K328-F437 G-protein coupled receptor BL00237:W367- F406_F478-Y489_R514-F540 Rhodopsin-like GPCR superfamily signature PR00237:R381-I403_T415-Y436_F470-S493 E519-V543_F561-R587_V301-F325_T334-F355 Neuropeptide Y receptor PRO101:R326-I356 Transmembrane domain: 7 transmembrane domain transmem_ domain: T302-F320 7 transmembrane receptor (rhodopsin family) 7tm_1:G316-N575 G-Protein Receptor V387-I403 G-protein coupled receptors signature S-protein_receptor.prf.N378-I428	BLAST_DOMO BLAST_PRODUM BLIMPS_BLOCKS BLIMPS_PRINTS HMMER HMMER_PPFAM MOTIFS PROFILESSCAN

WO 02/10387

PCT/US01/23433

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 Software
13	7482007CD1	695	S185 S252 S354 N169 N277 S469 T365 T337 T45 T584 T624 T653	N169 N277 N250 N257 N263 N286 N309 N340 N379 N61 N679 N686	HORMONE: EMRL; LEUCOCYTB; ANTIGEN; DM05221 L37251 347-738; C349-S688 RECEPTOR TRANSMEMBRANE G-PROTEIN COUPLED GLYCOPROTEIN PRECURSOR SIGNAL TYPE POLYPEPTIDE ALTERNATIVE PD000752:H372-R650 G-protein coupled receptors family 2 proteins. BL00649:C473-I498 Secretin-like GPCR superfamily signature PR00249:Y403-W427, A475-I498 Transmembrane domain transmem_domain:I407-T433, S514-P537, L467-F506, L552-P573 G-protein coupled receptor (secretin family) Tm 2: D398-I659	BLAST_DOMO BLAST_PRODUM BLIMPS_BLOCKS BLIMPS_PRINTS HMMER
14	5765042CD1	533	S108 S262 S275 S381 S589 S612 S616 S70 T251 T355 T517	N153 N235 N260 N263 N252 N366 N41 N604 N61 N78	HORMONE: EMRL; LEUCOCYTB; ANTIGEN; DM05221 A57174 465-886; G279-S593 RECEPTOR TRANSMEMBRANE G-PROTEIN COUPLED GLYCOPROTEIN PRECURSOR SIGNAL TYPE POLYPEPTIDE ALTERNATIVE PD000752:H21-K579 G-protein coupled receptors family 2 proteins. BL00649: S33-L360, C39-L446 G-protein coupled receptor (secretin family) Tm 2: D398-I659 Secretin-like GPCR superfamily signature PR00249: S327-S351, V393-L416, H431-8456, W473-8498, T517-L537, Q547-L568 Transmembrane domain transmem_domain: C333-L350, I472-T494 7 transmembrane receptor (Secretin family) Tm 2: Q322-V572 Lactoplin/1-1-like GFS domain C345-L365 Secretin coupled receptors family 2 signatures g_protein_recep_f2_2.psf: R474- S593	BLAST_DOMO BLAST_PRODUM BLIMPS_BLOCKS BLIMPS_PRINTS BLIMPS_PRINTS HMMER HMMER_FFAM HMMER_FFAM PROFLIBSCAN

WO 02/10387

PCT/US01/23433

Table 3 (cont.)

SEQ ID NO: 17	Incyte Polypeptide ID: 35036418CD1	Amino Acid Residues: 115	Potential Phosphorylation Sites: 568 588 588	Potential Glycosylation Sites: Mg	Signature Sequences, Domains and Motifs: Signal peptides: M1-R24 Transmembrane domains: L26-L46, C98-M119, E201-A320 7 transmembrane receptor (rhodopsin family): G42-Y291 G-protein coupled receptor motif: S111-I127 L104-G147 G-protein coupled receptors signature: R2L-F129, I108-F140, K23L-F243, R23L-F243, V29-V54, M60-K81, F105-I127, L200-L223, K273-M399 Olfactory receptor signature PR00245: M60-K81, F178-D192, L239-G254, A275-L286, S292-L306 G-protein coupled receptor: DM00013 P23275 I17-306: I18-I302 K3709 I5-304: V37-K304 P23270 I18-311: L26-L306 Olfactory G-protein coupled receptor: PD000921: Y169-L246 Olfactory G-protein coupled receptor: PDI49621: T247-R308	Analytical Methods and Databases: SSP, HMER, HMER-PFAM, MOTIFS, ProfileScan, BLIMPS-BLOCKS, BLIMPS-PRINTS, BLIMPS-PRINTS

WO 02/10387

PCT/US01/23433

Table 3 (cont.)

SEQ ID NO: 18	Incyte Polypeptide ID: 7481701CD1	Amino Acid Residues: 324	Potential Phosphorylation Sites: S257, S314, S370, S67, S87, T136, T8	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs: Transmembrane domains: L33-N52, V243-F263, V267-N287; 7 transmembrane receptor (rhodopsin family): G41-F180, V225-V291; Visual pigments (opsins) retinal binding site: W271-F319; G-protein coupled receptor: H000237; R33-F128, K135-K62, S282-K237; R33-F128, K135-K62, S282-K237; Family: PR00237; Y26-Q56, M59-K80, F103-I125, F12-F33, T144-F167, A238-R262, I273-K299; Olfactory receptor signature: PR00245; M59-K80, I176-D190, F239-G254, S292-I306; Melanocortin receptor family: PR00534; L51-I62, I125-T136; G-protein coupled receptor: DM00013; E23267/20-309: F17-I306; S23709/11-239: F29-K304; S23710/11-239: F29-K304; E23274/18-306: V26-I302; Olfactory G-protein coupled receptor: PD000921; L185-I247	Analytical Methods and Databases: HMHR, HMHR-PFAM, ProfilesScan, BLIMPS-BLOCKS, BLIMPS-PRINTS, BLIMPS-PRINTS, BLIMPS-PRINTS, BLAST-DBOM, BLAST-PRODOM

WO 02/10387

PCT/US01/23433

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 Software
19	748177ACD1	312	S16 S230 S264 S55 I161 T289	N40	Transmembrane domains: I24-V46, Q98-M116, F189-Y211 7 transmembrane receptor (rhodopsin family): G39-Y288 G-protein coupled receptor motif: V108-I124 G-protein coupled receptors signature: F100-T146 G-protein coupled receptor BL00237: F88-F127, Y205-Y216, Q233-Q259, T280-K296 Olfactory receptor signature PRO0243: K37-Q79, F173-S189, F236-G251, L272-L283, R282-G312 G-protein coupled receptor: DM00013 P23275117-306: S16-L303 A5706915-304: F15-L303 P3498217-305: S16-L303 P3093318-306: R18-L303 Olfactory G-protein coupled receptor: ED145621: T244-R307 Olfactory G-protein coupled receptor: PD000921: C187-S243	HMMER HMMER-PTAM MOTIFS ProfileScan BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-DOMO BLAST-PRODOM BLAST-PRODOM

WO 02/10387

PCT/US01/23433

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
20	747480CB1	1076	1-1076	7075196HL (BRD1ERR04)	1	532
21	7474840CB1	1102	1007-1102, 107-160, 486-522, 642-842	g7407827_edit	533	1102
22	7475092CB1	2529	628-1475, 1-81	55049853HL 73906055V1 13518366E (LAFRTVTF02) GRR137023955_000031_002 8079820V1 5079820V1 (BRARFYE01) 55049853HL 55049853HL	876 564 1757 1988 64 64 1956	1612 209 267 469 639 644 1956
23	7341260CB1	1847	1-140, 1490-1847	7341260HL (COLNDIN02) 70811587V1 683479HL (BRSTNON02) 70888652V1 7628842HL (GELADIE01)	307 1239 974 228	991 1847 1575 660
24	7473911CB1	2031	1-504, 669-1834	FL140044_00001	1	2031
25	7474767CB1	1130	1-1130	GNN_66693326_000106_002	139	1130
26	7475815CB1	1202	357-949, 1044-1202, 1-1165	5079820V1 52747815_68197595_000004_372 63882596	1 372	1130 723 1202
27	60263275CB1	2079	134-430, 902-981	71704087V1 71651942V1 3642425NG (LUNGNGT34) 524802R6 (CARCTXT01) 2435123HL (BRAYVNF02) 71651560V1	85 986 1469 391 1 838	700 1727 2079 985 255 1544

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
28	6020310CB1	5324	1-373-661, 263-5661, 4111-4162	7195931V1	4179	5068
				6020311D1	329	562
				6020312D1 (BRATFER01)	333	4073
				4194941 (HNZ2ACT01)	343	4073
				7195769V1	4656	5324
				4028716V1 (BRATNC023)	4057	4787
				8103587H1 (MEXDIE02)	620	1259
				9800324	3652	4165
				8195081H2 (BRAIDIR04)	1	623
				5507339Q1	1549	2391
				6020067D1	3444	3823
				8195081J2 (BRAIDIR04)	432	1033
				5509409J1	1140	2026
29	7477349CB1	1362	1-1758	5116120H1	2658	3416
				7084648V1	1488	1962
				4194941 (HNZ2ACT01)	165	379
				2021568P6 (CONNNGT01)	379	873
				7124343V1	1272	1931
				7084423V1	872	1416
				7084576V1	804	1415
				7239821V1	675	1384
				7237437V1	971	1558
				72373094V1	814	1540
				5509494J1	1	817
				CGS_SAE300482544_R1	1148	1481
				GNN:G6138786_000002_006_ed	1	1893
30	55002225CB1	1558	80-1558, 1-49	7590465P6 (BRATFER06)	1576	2304
				5509494J1	461	1377
				6925371H1 (EACFER06)	249	797
				GNN:G9864547_000007_002_ed	472	2103
				.it		
31	7475886CB1	2304	1-1212, 1548-2304	55084155J2	1017	3790
				7341368F8 (COLINDING02)	1	440
				91507289	1879	2322
32	7482007CB1	2322	1-628, 1452-1878, 724-1278			

WO 02/10387

PCT/US01/23433

Table 4

Polynucleotide SEQ ID NO.	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
33	6769042CB1	2366	800-1248, 1-40, 1717, 2183-264, 1831-1861, 1930-2366, 1-1087	7157074E2 (BRA1FE02) 72138116D1 7629277H1 (GEGADIER01) 5514760601	1169 1704 834 1	1797 2466 1341 908
34	7476053CB1	1458	1-1087	GNN.G7630808.000013_022 G2280262	1201 1088	1414 1458
35	7480410CB1	975	1-816, 894- 975	GRI.G9454621.edit 55036194H1 (GPCRPFY02)	1 166	1110 371
36	55036418CB1	948	1-162, 193- 246, 1086	GNN.G6979559.000011_002 55036393E1 (GPCRPFY02)	1 166	975 371
37	7481701CB1	1086	1-1086	GNN.G7232840.000072_008	1	1086
38	7481774CB1	1523	1-983	5514353571 7082263V1	127 879	918 1529
				7361408F8 (BRA1FE05) 55142634H1	401 1	1051 384

WO 02/10387

PCT/US01/23433

Table 5

Polynucleotide SEQ. ID. NO.	Incyte Project ID	Representative Library
20	7474808CE1	BRAUTDR04
22	7475092CE1	LATRTUT02
23	7341260CE1	COLATUT03
24	7473911CE1	BESTMGT03
27	60263275CE1	EOSFTXT01
28	60203310CE1	BRAUTVT01
29	7477349CE1	COMMOT01
31	7475686CE1	BRAIFER06
32	7482007CE1	COLADIN02
33	6769042CE1	GLADIE01
38	7481774CE1	BRAIFEE05

WO 02/10387

PCT/US01/23433

Table 6

Library	Vector	Library Description
BRA1FEB05	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from brain tissue removed from Caucasian male fetus who was still born with a hypoplastic left heart at 23 weeks gestation. Serologies were negative.
BRA1FER06	PCDNA2.1	This random primed library was constructed using RNA isolated from brain tissue removed from a 35-year-old Caucasian male fetus who stillborn with a hypoplastic left heart at 23 weeks gestation. Serologies were negative.
BRA1UT01	PSFOET1	Library was constructed using RNA isolated from brain tissue removed from a 50-year-old Caucasian female during a frontal lobectomy. Pathology indicated recurrent grade 3 oligoastrocytoma with focal necrosis and extensive calcification. Patient history included a speech disturbance and epilepsy. The patient's brain had also been irradiated with a total dose of 5,082 cGy (fraction 8). Family history included a brain tumor.
BRAUTDR04	PCDNA2.1	This random primed library was constructed using RNA isolated from pooled striatum tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma. Postoperative medical treatments included chemotherapy, radiation, and surgical resection of the liver and extrahepatic biliary system. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRSTN0T23	PINCY	Library was constructed using RNA isolated from diseased breast tissue removed from a 35-year-old Caucasian female during a bilateral reduction mammoplasty. Pathology indicated nonproliferative fibrocystic disease. Family history included type II diabetes, atherosclerotic coronary artery disease, acute myocardial infarction, hyperlipidemia, and coronary artery bypass.
COLANDIN02	PINCY	This normalized library was constructed from 4.72 million independent clones from a diseased colon ampullon polyp tissue library. The library was made from polyp tissue removed from the cecum and descending colon of a 16-year-old Caucasian male (donor A) during partial colectomy, temporary ileostomy, and colonoscopy and from diseased colon polyp tissue removed from the cecum of a 67-year-old female (donor B). Pathology indicated innumerable (greater than 100) adenomatous polyps with low-grade dysplasia involving the entire colonic mucosa in the setting of familial polyposis coli (donor A), and a benign cecum polyp (donor B). Pathology for the associated tumor tissue (B) indicated massive proctitis and adenocarcinoma. The adenocarcinoma was characterized by multiple (2 of 17) regional lymph nodes were involved by metastatic adenocarcinoma. Multiple (3 of 17) regional lymph nodes were involved by metastatic adenocarcinoma. A tubulovillous adenoma and multiple (6) tubular adenomas with low-grade dysplasia were observed in the cecum and ascending colon. Donor A presented with abdominal pain and

WO 02/10387

PCT/US01/23433

Table 6 (cont.)

Library	Vector	Library Description
COLMFWU03	P1N1CY	Library Description filantule. The patient was not taking any medications. Family history included benign colon neoplasm in the father and sibling(s); benign hypertension, cerebrovascular disease, breast cancer, uterine cancer, and Type II diabetes in the grandparent(s). Library was constructed using RNA isolated from colon tumor tissue obtained from the sigmoid colon of a 62-year-old Caucasian male. The patient's medical history included metastasis with extranodal extension. Patient history included hyperlipidemia, cataract disorder, and dermatitis. Family history included benign hypertension, atherosclerotic coronary artery disease, hyperlipidemia, breast cancer, and prostate cancer.
COMMNW01	P1N1CY	Library was constructed using RNA isolated from mesentery fat tissue obtained from a 71-year-old Caucasian male during a partial colectomy and permanent colostomy. Family history included atherosclerotic coronary artery disease, myocardial infarction, and extrinsic asthma.
ESSTFW01	P1N1CY	Library was constructed using RNA isolated from scapulothorax stimulated with II-5
GBLADLE01	PCMR2.1	Library was constructed using RNA isolated from a 55-year-old Caucasian female with diseased gallbladder. Pathology indicated chronic cholecystitis and cholelithiasis (greater than 100 stones). The patient presented with cholelithiasis, abdominal pain, and tremors. Patient history included benign hypertension, Morton's neuroma, facial hirsutism, normal delivery, and tobacco abuse in remission. Previous surgeries included total abdominal hysterectomy, bilateral salpingo-oophorectomy, and adenotonsillectomy. Patient medications included indinavir and Protonix. Family history included cancer in the grandparent(s) and atherosclerotic coronary artery disease in the grandparent(s).
LAFRWU02	P1N1CY	Library was constructed using RNA isolated from a myxoma removed from the left atrium of a 43-year-old Caucasian male during annuloplasty. Pathology indicated atrial myxoma. Patient history included pulmonary insufficiency, acute myocardial infarction, atherosclerotic coronary artery disease, hyperlipidemia, and tobacco use. Family history included benign hypertension, acute myocardial infarction, atherosclerotic coronary artery disease, and Type II diabetes.

WO 02/10387

PCT/US01/23433

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 PARACIL PDF	A Fast Data Finder useful in computing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) <i>J. Mol. Biol.</i> 215:403-410; Altschul, S.F. et al. (1997) <i>Nucleic Acids Res.</i> 25:3389-3402.	ESY: Probability value= 1.0E-8 or less Full Length 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, ifasta, fastx, ifastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) <i>Proc. Natl. Acad. Sci. USA</i> 85:2444-2448; Pearson, W.R. (1990) <i>Methods Enzymol.</i> 183:63-98; and Smith, T.F. and M.S. Waterman (1981) <i>Adv. Appl. Math.</i> 2:482-489.	ESY: fasta E value=1.0E-6 Assembled ESY: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx scores=100 or greater
BLIMFS	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 Attwood, T.K. et al. (1997) <i>J. Chem. Inf. Comput. Sci.</i> 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) <i>J. Mol. Biol.</i> 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) <i>Nucleic Acids Res.</i> 26:320-322; Durbin, R. et al. (1998) <i>Our World View</i> , in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less Signal peptide hits: Scores= 0 or greater

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:GCC-specified "HIGH" value for that particular Prosite motif. Generally, score=1,4-2,1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A 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.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Clavette, J.M. and S. Audic (1997) CABIOS 12:431-439.	Scores=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 Int. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

WO 02/10387

PCT/US01/23433

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:1-19,
 - 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:1-19,
 - c) a biologically active fragment of a polypeptide having an amino acid sequence selected
from the group consisting of SEQ ID NO:1-19, and
 - 10 d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from
the group consisting of SEQ ID NO:1-19.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-
19.
- 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:20-38.
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 02/10387

PCT/US01/23433

- 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:20-38,
- 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:20-38,
- 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 02/10387

PCT/US01/23433

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
5 from the group consisting of SEQ ID NO:1-19.
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.
10
19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:
a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
b) detecting agonist activity in the sample.
15
20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.
21. A method for treating a disease or condition associated with decreased expression of
20 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:
25 a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
b) detecting antagonist activity in the sample.
23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.
30
24. A method for treating a disease or condition associated with overexpression of functional 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

WO 02/10387

PCT/US01/23433

1, said method comprising the steps of:

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

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

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

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

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

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

- a) treating a biological sample containing nucleic acids with the test compound;
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;
- c) quantifying the amount of hybridization complex; and
- d) comparing the amount of hybridization complex in the treated biological sample with the

WO 02/10387

PCT/US01/23433

amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

5 29. A diagnostic test for a condition or disease associated with the expression of 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

10 b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

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

15 a) a chimeric antibody,
b) a single chain antibody,

c) a Fab fragment,

d) a F(ab')₂ fragment, or

e) a humanized antibody.

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

32. A method of diagnosing a condition or disease associated with the expression of 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:1-19, or an immunogenic fragment thereof, under conditions to

WO 02/10387

PCT/US01/23433

elicit an antibody response;

- b) isolating antibodies from said animal; and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19.

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

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

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:1-19, or an immunogenic fragment thereof, under conditions to elicit an antibody response;
- b) isolating antibody producing cells from the animal;
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;
- d) culturing the hybridoma cells; and
- e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19.

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

43. A method for detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 in a sample, comprising the steps of:

WO 02/10387

PCT/US01/23433

- a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and
- b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 in the sample.

44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 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
- 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:1-19.

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

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

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

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

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.

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.

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 02/10387

PCT/US01/23433

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.
- 5 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.
- 10 61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.
62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.
63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.
- 15 64. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:20.
- 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.
- 25 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.
- 30 69. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:25.

WO 02/10387

PCT/US01/23433

70. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:26.
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 NO:28.
73. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:29.
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.
76. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:32.
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.
79. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:35.
80. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:36.
81. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID

WO 02/10387

PCT/US01/23433

NO:37.

82. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:38.

WO 02/10387

PCT/US01/23433

<110> INCYTE GENOMICS, INC.
 THORNTON, Michael
 PATTERSON, Chandra
 LAL, Preeti
 BURFORD, Neil
 YUE, Henry
 GANDHI, Ameena R.
 ELLIOTT, Vicki S.
 RAMKUMAR, Jayalaxmi
 BAUGHN, Mariah R.
 KALLICK, Deborah A.
 WALIA, Narinder K.
 HAPALIA, April J.A.
 YAO, Monique G.
 LU, Yan
 TRIBOULEY, Catherine M.
 POLICKY, Jennifer L.
 KEARNEY, Liam
 GRAUL, Richard
 WARREN, Bridget
 LEE, Ernestine A.
 DING, Li

<120> G-PROTEIN COUPLED RECEPTORS

<130> PI-0176 PCT

<140> To Be Assigned
 <141> Herewith

<150> 60/221,478; 60/223,268; 60/227,054; 60/231,121; 60/232,243;
 60/232,691; 60/235,146
 151> 2000-07-27; 2000-08-03; 2000-08-21; 2000-09-08; 2000-09-13;
 2000-09-15; 2000-09-22

<160> 38
 <170> PERL Program

<210> 1
 <211> 339
 <212> FRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7474806CD1

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

WO 02/10387

PCT/US01/23433

```

110
Leu Leu Cys Lys Leu Val His Phe Leu Phe Tyr Ile Asn Leu Tyr 120
125
Gly Ser Ile Leu Leu Leu Thr Cys Ile Ser Val His Gln Phe Leu 135
140
Gly Val Trp His Pro Leu Cys Ser Leu Pro Tyr Arg Thr Arg Arg 150
155
His Ala Trp Leu Gly Thr Ser Thr Thr Trp Ala Leu Val Val Leu 165
170
Gln Leu Leu Pro Thr Leu Ala Phe Ser His Thr Asp Tyr Ile Asn 180
185
Gly Gln Met Ile Trp Tyr Asp Met Thr Ser Gln Glu Asn Phe Asp 195
200
Arg Leu Phe Ala Tyr Gly Ile Val Leu Thr Leu Ser Gly Phe Leu 210
215
Ser Pro Ser Leu Val Ile Leu Val Cys Tyr Ser Leu Met Val Arg 225
230
Ser Leu Ile Lys Pro Glu Glu Asn Leu Met Arg Thr Gly Asn Thr 240
245
Ala Arg Ala Arg Ser Ile Arg Thr Ile Leu Leu Val Cys Gly Leu 255
260
Phe Thr Leu Cys Phe Val Pro Phe His Ile Thr Arg Ser Phe Tyr 270
275
Leu Thr Ile Cys Phe Leu Leu Ser Gln Asp Cys Gln Leu Leu Met 285
290
Ala Pro Ser Val Ala Tyr Lys Ile Trp Arg Pro Leu Val Ser Val 300
305
Ser Ser Cys Leu Asn Pro Val Leu Tyr Phe Leu Ser Arg Gly Ala 315
320
Lys Ile Glu Ser Gly Ser Ser Arg Asn 325
335

```

```

<210> 2
<211> 335
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7474840CD1

```

```

<400> 2
Met Thr Pro Gly Gly Arg Ala Cys Ser Glu Met Arg Ser Cys His
1 5 10 15
Cys Ala Pro Ala Trp Ala Thr Glu Arg Asp Ser Val Ser Lys Lys
20 25 30
Lys Lys Asn Lys Lys Lys Asn Leu Phe Ser Gln Ala Thr Ile Gly
35 40 45
Leu Leu Ala Asn Thr Phe Phe Leu Phe Phe Asn Ile Phe Ile Phe
50 55 60
Leu Gln Asp Gln Lys Ser Lys Pro His Asp Leu Ile Ser Cys Asn
65 70 75
Ser Ala Phe Ile His Val Val Met Phe Leu Thr Val Val Asp Ala
80 85 90
Trp Pro Pro Asp Met Pro Glu Ser Leu His Leu Gly Asn Glu Phe
95 100 105
Lys Phe Lys Ser Leu Ser Tyr Ile Asn Arg Val Arg Met Gly Leu
110 115 120
Cys Ile Cys Asn Ile Cys Leu Leu Ser Ile His Gln Ala Asn Thr
125 130 135
Ile Ser Pro Asn Asn Phe Cys Leu Ala Arg Leu Lys Gln Lys Phe
140 145 150
Thr Asn Asn Ile Ile Met Ser Ser Phe Phe Ser Phe Phe Phe Trp

```

WO 02/10387

PCT/US01/23433

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

<210> 3
 <211> 428
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> misc_feature
 <223> Incyte ID No: 7475092CD1

<400> 3
 Met Gln Arg Lys Glu Lys Ala Lys Cys Pro Gln Glu Ala Pro Ala
 1 5 10 15
 Gly Arg Glu Pro Ser Thr Pro Gly Gly Gly Ser Gly Gly Gly Gly
 20 25 30
 Ala Val Ala Ala Ala Ser Gly Ala Ala Val Pro Gly Ser Val Gln
 35 40 45
 Leu Ala Leu Ser Val Leu His Ala Leu Leu Tyr Ala Ala Leu Phe
 50 55 60
 Ala Phe Ala Tyr Leu Gln Leu Trp Arg Leu Leu Leu Tyr Arg Glu
 65 70 75
 Arg Arg Leu Ser Tyr Gln Ser Leu Cys Leu Phe Leu Cys Leu Leu
 80 85 90
 Trp Ala Ala Leu Arg Thr Thr Leu Phe Ser Ala Ala Phe Ser Leu
 95 100 105
 Ser Gly Ser Leu Pro Leu Leu Arg Pro Pro Ala His Leu His Phe
 110 115 120
 Phe Pro His Trp Leu Leu Tyr Cys Phe Pro Ser Cys Leu Gln Phe
 125 130 135
 Ser Thr Leu Cys Leu Leu Asn Leu Tyr Leu Ala Glu Val Ile Cys
 140 145 150
 Lys Val Arg Cys Ala Thr Glu Leu Asp Arg His Lys Ile Leu Leu
 155 160 165
 His Leu Gly Phe Ile Met Ala Ser Leu Leu Phe Leu Val Val Asn
 170 175 180
 Leu Thr Cys Ala Met Leu Val His Gly Asp Val Pro Glu Asn Gln
 185 190 195
 Leu Lys Trp Thr Val Phe Val Arg Ala Leu Ile Asn Asp Ser Leu

WO 02/10387

PCT/US01/23433

200 205 210
 Phe Ile Leu Cys Ala Ile Ser Leu Val Cys Tyr Ile Cys Lys Ile
 215 220 225
 Thr Lys Met Ser Ser Ala Asn Val Tyr Leu Glu Ser Lys Gly Met
 230 235 240
 Ser Leu Cys Gln Thr Val Val Val Gly Ser Val Val Ile Leu Leu
 245 250 255
 Tyr Ser Ser Arg Ala Cys Tyr Asn Leu Val Val Val Thr Ile Ser
 260 265 270
 Gln Asp Thr Leu Glu Ser Pro Phe Asn Tyr Gly Trp Asp Asn Leu
 275 280 285
 Ser Asp Lys Ala His Val Glu Asp Ile Ser Gly Glu Glu Tyr Ile
 290 295 300
 Val Phe Gly Met Val Leu Phe Leu Trp Glu His Val Pro Ala Trp
 305 310 315
 Ser Val Val Leu Phe Phe Arg Ala Gln Arg Leu Asn Gln Asn Leu
 320 325 330
 Ala Pro Ala Gly Met Ile Asn Ser His Ser Tyr Ser Ser Arg Ala
 335 340 345
 Tyr Phe Phe Asp Asn Pro Arg Arg Tyr Asp Ser Asp Asp Asp Leu
 350 355 360
 Pro Arg Leu Gly Ser Ser Arg Glu Gly Ser Leu Pro Asn Ser Gln
 365 370 375
 Ser Leu Gly Trp Tyr Gly Thr Met Thr Gly Cys Gly Ser Ser Ser
 380 385 390
 Tyr Thr Val Thr Pro His Leu Asn Gly Pro Met Thr Asp Thr Ala
 395 400 405
 Pro Leu Leu Phe Thr Cys Ser Asn Leu Asp Leu Asn Asn His His
 410 415 420
 Ser Leu Tyr Val Thr Pro Gln Asn
 425

<210> 4
 <211> 330
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7341260CD1

<400> 4
 Met Thr Pro Asn Ser Thr Gly Glu Val Pro Ser Pro Ile Pro Lys
 1 5 10 15
 Gly Ala Leu Gly Leu Ser Leu Ala Leu Ala Ser Leu Ile Ile Thr
 20 25 30
 Ala Asn Leu Leu Leu Ala Leu Gly Ile Ala Trp Asp Arg Arg Leu
 35 40 45
 Arg Ser Pro Pro Ala Gly Cys Phe Phe Leu Ser Leu Leu Leu Ala
 50 55 60
 Gly Leu Leu Thr Gly Leu Ala Leu Pro Thr Leu Pro Gly Leu Trp
 65 70 75
 Asn Gln Ser Arg Arg Gly Tyr Trp Ser Cys Leu Leu Val Tyr Leu
 80 85 90
 Ala Pro Asn Phe Ser Phe Leu Ser Leu Leu Ala Asn Leu Leu Leu
 95 100 105
 Val His Gly Glu Arg Tyr Met Ala Val Leu Arg Pro Leu Gln Pro
 110 115 120
 Pro Gly Ser Ile Arg Leu Ala Leu Leu Leu Thr Trp Ala Gly Pro
 125 130 135
 Leu Leu Phe Ala Ser Leu Pro Ala Leu Gly Trp Asn His Trp Thr
 140 145 150
 Pro Gly Ala Asn Cys Ser Ser Gln Ala Ile Phe Pro Ala Pro Tyr

WO 02/10387

PCT/US01/23433

```

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

```

```

<210> 5
<211> 676
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473911CD1

```

```

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

```

WO 02/10387

PCT/US01/23433

Lys Cys Ser Glu Ser Glu Thr Asp Glu Asp Tyr Ile Val Leu Arg
215 220 225
Ala Pro Leu Arg Glu Asp Glu Pro Lys Asp Gly Gly Ser Val Gly
230 235 240
Asn Ala Ala Leu Val Ser Pro Glu Ala Ser Ala Glu Glu Glu Glu
245 250 255
Glu Arg Glu Glu Gly Gly Glu Ala Cys Gly Leu Glu Arg Thr Gly
260 265 270
Ala Gly Gly Glu Gln Val Asp Leu Gly Glu Leu Pro Asp His Glu
275 280 285
Glu Lys Ser Asn Gln Lys Val Ala Ala Ala Thr Leu Glu Asp Arg
290 295 300
Thr Gln Asp Glu Pro Ala Glu Glu Ser Cys Gln Ile Val Leu Phe
305 310 315
Gln Asn Asn Cys Met Asp Asn Phe Val Thr Ser Leu Thr Gly Ser
320 325 330
Pro Tyr Glu Phe Phe Pro Thr Lys Ser Thr Ser Phe Cys Arg Glu
335 340 345
Ser Cys Ser Pro Phe Ser Glu Ser Val Lys Ser Leu Glu Ser Glu
350 355 360
Gln Ala Pro Lys Leu Gly Leu Cys Ala Glu Glu Asp Pro Val Val
365 370 375
Gly Ala Leu Cys Gly Gln His Gly Pro Leu Gln Asp Gly Val Ala
380 385 390
Glu Gly Pro Thr Ala Pro Asp Val Val Val Leu Pro Lys Glu Glu
395 400 405
Glu Lys Glu Glu Val Ile Val Asp Asp Met Leu Ala Asn Pro Tyr
410 415 420
Val Met Gly Asp Glu Gly Glu Glu Glu Glu Glu Phe Val Asp
425 430 435
Asp Thr Leu Ala Asn Pro Tyr Val Met Gly Val Gly Leu Pro Gly
440 445 450
Arg Gly Gly Glu Glu Glu Glu Glu Glu Val Val Asp Asp Thr
455 460 465
Leu Ala Ser Leu Tyr Lys Met Gly Glu Glu His Arg His Lys Gly
470 475 480
Leu Ala Pro Leu Trp Glu Gly Gly Gln Lys Pro Ser Gln Lys Leu
485 490 495
Pro Pro Lys Lys Pro Asp Leu Arg Gln Val Pro Gln Pro Leu Ala
500 505 510
Ser Glu Val Pro Gln Arg Arg Gln Glu Arg Ala Val Val Thr Glu
515 520 525
Gly Arg Pro Leu Glu Ala Ser Arg Ala Leu Pro Ala Lys Pro Arg
530 535 540
Ala Phe Thr Leu Tyr Pro Arg Ser Phe Ser Val Glu Gly Gln Glu
545 550 555
Ile Pro Val Ser Ile Ser Val Tyr Trp Glu Pro Glu Gly Ser Gly
560 565 570
Leu Asp Asp His Arg Ile Lys Arg Lys Glu Glu His Leu Ser Val
575 580 585
Val Ser Gly Ser Phe Ser Gln Arg Asn His Leu Pro Ser Ser Gly
590 595 600
Thr Ser Thr Pro Ser Ser Met Val Asp Ile Pro Pro Pro Phe Asp
605 610 615
Leu Ala Cys Ile Thr Lys Lys Pro Ile Thr Lys Ser Ser Pro Ser
620 625 630
Leu Leu Ile Asp Ser Asp Ser Pro Asp Lys Tyr Lys Lys Lys Lys
635 640 645
Ser Ser Phe Lys Arg Phe Leu Ala Leu Met Phe Asn Lys Met Glu
650 655 660
Arg Pro Gly Thr Met Ala His Ala Cys His Pro Ser Thr Leu Gly
665 670 675
Ser

WO 02/10387

PCT/US01/23433

```

<210> 6
<211> 372
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474767CD1

<400> 6
Met Glu His Thr His Ala His Leu Ala Ala Asn Ser Ser Leu Ser
 1          5          10          15
Trp Trp Ser Pro Gly Ser Ala Cys Gly Leu Gly Phe Val Pro Val
 20          25          30
Val Tyr Tyr Ser Leu Leu Leu Cys Leu Gly Leu Pro Ala Asn Ile
 35          40          45
Leu Thr Val Ile Ile Leu Ser Gln Leu Val Ala Arg Arg Gln Lys
 50          55          60
Ser Ser Tyr Asn Tyr Leu Leu Ala Leu Ala Ala Asp Ile Leu
 65          70          75
Val Leu Phe Phe Ile Val Phe Val Asp Phe Leu Leu Glu Asp Phe
 80          85          90
Ile Leu Asn Met Gln Met Pro Gln Val Pro Asp Lys Ile Ile Glu
 95          100          105
Val Leu Glu Phe Ser Ser Ile His Thr Ser Ile Trp Ile Thr Val
 110          115          120
Pro Leu Thr Ile Asp Arg Tyr Ile Ala Val Cys His Pro Leu Lys
 125          130          135
Tyr His Thr Val Ser Tyr Pro Ala Arg Thr Arg Lys Val Ile Val
 140          145          150
Ser Val Tyr Ile Thr Cys Phe Leu Thr Ser Ile Pro Tyr Tyr Trp
 155          160          165
Trp Pro Asn Ile Trp Thr Glu Asp Tyr Ile Ser Thr Ser Val His
 170          175          180
His Val Leu Ile Trp Ile His Cys Phe Thr Val Tyr Leu Val Pro
 185          190          195
Cys Ser Ile Phe Phe Ile Leu Asn Ser Ile Ile Val Tyr Lys Leu
 200          205          210
Arg Arg Lys Ser Asn Phe Arg Leu Arg Gly Tyr Ser Thr Gly Lys
 215          220          225
Thr Thr Ala Ile Leu Phe Thr Ile Thr Ser Ile Phe Ala Thr Leu
 230          235          240
Trp Ala Pro Arg Ile Ile Met Ile Leu Tyr His Leu Tyr Gly Ala
 245          250          255
Pro Ile Gln Asn Arg Trp Leu Val His Ile Met Ser Asp Ile Ala
 260          265          270
Asn Met Leu Ala Leu Leu Asn Thr Ala Ile Asn Phe Phe Leu Tyr
 275          280          285
Cys Phe Ile Ser Lys Arg Phe Arg Thr Met Ala Ala Ala Thr Leu
 290          295          300
Lys Ala Phe Phe Lys Cys Gln Lys Gln Pro Val Gln Phe Tyr Thr
 305          310          315
Asn His Asn Phe Ser Ile Thr Ser Ser Pro Trp Ile Ser Pro Ala
 320          325          330
Asn Ser His Cys Ile Lys Met Leu Val Tyr Gln Tyr Asp Lys Asn
 335          340          345
Gly Lys Pro Ile Lys Ser Arg Asn Asp Ser Lys Ser Ser Tyr Gln
 350          355          360
Phe Glu Asp Ala Ile Gly Ala Cys Val Ile Ile Leu
 365          370

```

WO 02/10387

PCT/US01/23433

<210> 7
<211> 271
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475815CD1

```

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

```

<210> 8
<211> 611
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 60263275CD1

```

<400> 8
Met Gln Gly Pro Leu Leu Leu Pro Gly Leu Cys Phe Leu Leu Ser
1 5 10 15
Leu Phe Gly Ala Val Thr Gln Lys Thr Lys Asn Ile Asn Glu Cys
20 25 30

```

WO 02/10387

PCT/US01/23433

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

WO 02/10387

PCT/US01/23433

500
 Phe Lys Ala Thr Ala Gln Leu Phe Ile Leu Gly Cys Thr Trp Cys 510
 515
 Leu Gly Leu Leu Gln Val Gly Pro Ala Ala Gln Val Met Ala Tyr 525
 530
 Leu Phe Thr Ile Ile Asn Ser Leu Gln Gly Phe Phe Ile Phe Leu 540
 545
 Val Tyr Cys Leu Leu Ser Gln Gln Val Gln Lys Gln Tyr Gln Lys 555
 560
 Trp Phe Arg Glu Ile Val Lys Ser Lys Ser Glu Ser Glu Thr Tyr 570
 575
 Thr Leu Ser Ser Lys Met Gly Pro Asp Ser Lys Pro Ser Glu Gly 585
 590
 Asp Val Phe Pro Gly Gln Val Lys Arg Lys Tyr 600
 605

<210> 9
 <211> 3469
 <212> PRT
 <213> Homo sapiens

 <220>
 <221> misc_feature
 <223> Incyte ID No: 60203310CD1

<400> 9
 Met Trp Pro Ser Gln Leu Leu Ile Phe Met Met Leu Leu Ala Pro 1
 1 5 10 15
 Ile Ile His Ala Phe Ser Arg Ala Pro Ile Pro Met Ala Val Val 20
 20 25 30
 Arg Arg Glu Leu Ser Cys Glu Ser Tyr Pro Ile Glu Leu Arg Cys 35
 35 40 45
 Pro Gly Thr Asp Val Ile Met Ile Glu Ser Ala Asn Tyr Gly Arg 50
 50 55 60
 Thr Asp Asp Lys Ile Cys Asp Ser Asp Pro Ala Gln Met Glu Asn 65
 65 70 75
 Ile Arg Cys Tyr Leu Pro Asp Ala Tyr Lys Ile Met Ser Gln Arg 80
 80 85 90
 Cys Asn Asn Arg Thr Gln Cys Ala Val Val Ala Gly Pro Asp Val 95
 95 100 105
 Phe Pro Asp Pro Cys Pro Gly Thr Tyr Lys Tyr Leu Glu Val Gln 110
 110 115 120
 Tyr Glu Cys Val Pro Tyr Lys Val Glu Gln Lys Val Phe Leu Cys 125
 125 130 135
 Pro Gly Leu Leu Lys Gly Val Tyr Gln Ser Glu His Leu Phe Glu 140
 140 145 150
 Ser Asp His Gln Ser Gly Ala Trp Cys Lys Asp Pro Leu Gln Ala 155
 155 160 165
 Ser Asp Lys Ile Tyr Tyr Met Pro Trp Thr Pro Tyr Arg Thr Asp 170
 170 175 180
 Thr Leu Thr Glu Tyr Ser Ser Lys Asp Asp Phe Ile Ala Gly Arg 185
 185 190 195
 Pro Thr Thr Thr Tyr Lys Leu Pro His Arg Val Asp Gly Thr Gly 200
 200 205 210
 Phe Val Val Tyr Asp Gly Ala Leu Phe Phe Asn Lys Glu Arg Thr 215
 215 220 225
 Arg Asn Ile Val Lys Phe Asp Leu Arg Thr Arg Ile Lys Ser Gly 230
 230 235 240
 Glu Ala Ile Ile Ala Asn Ala Asn Tyr His Asp Thr Ser Pro Tyr 245
 245 250 255
 Arg Trp Gly Gly Lys Ser Asp Ile Asp Leu Ala Val Asp Glu Asn 260
 260 265 270
 Gly Leu Trp Val Ile Tyr Ala Thr Glu Gln Asn Asn Gly Lys Ile

WO 02/10387

PCT/US01/23433

Val Ile Ser Gln	275	280	285
Leu Asn Pro Tyr Thr	280	Leu Arg Ile Glu Gly Thr	
290	295	300	
Trp Asp Thr Ala	Tyr Asp Lys Arg Ser	Ala Ser Asn Ala Phe Met	
305	310	315	
Ile Cys Gly Ile	Leu Tyr Val Val Lys Ser	Val Tyr Glu Asp Asp	
320	325	330	
Asp Asn Glu Ala	Thr Gly Asn Lys Ile Asp Tyr Ile Tyr Asn Thr		
335	340	345	
Asp Gln Ser Lys	Asp Ser Leu Val Asp Val Pro Phe Pro Asn Ser		
350	355	360	
Tyr Gln Tyr Ile	Ala Ala Val Asp Tyr Asn Pro Arg Asp Asn Leu		
365	370	375	
Leu Tyr Val Trp	Asn Asn Tyr His Val Val Lys Tyr Ser Leu Asp		
380	385	390	
Phe Gly Pro Leu	Asp Ser Arg Ser Gly Gln Ala His His Gly Gln		
395	400	405	
Val Ser Tyr Ile	Ser Pro Pro Ile His Leu Asp Ser Glu Leu Glu		
410	415	420	
Arg Pro Ser Val Lys	Asp Ile Ser Thr Thr Gly Pro Leu Gly Met		
425	430	435	
Gly Ser Thr Thr	Thr Ser Thr Thr Leu Arg Thr Thr Thr Leu Ser		
440	445	450	
Pro Gly Arg Ser	Thr Thr Pro Ser Val Ser Gly Arg Arg Asn Arg		
455	460	465	
Ser Thr Ser Thr	Pro Ser Pro Ala Val Glu Val Leu Asp Asp Met		
470	475	480	
Thr Thr His Leu	Pro Ser Ala Ser Ser Gln Ile Pro Ala Leu Glu		
485	490	495	
Glu Ser Cys Glu	Ala Val Glu Ala Arg Glu Ile Met Trp Phe Lys		
500	505	510	
Thr Arg Gln Gly	Gln Ile Ala Lys Gln Pro Cys Pro Ala Gly Thr		
515	520	525	
Ile Gly Val Ser	Thr Tyr Leu Cys Leu Ala Pro Asp Gly Ile Trp		
530	535	540	
Asp Pro Gln Gly	Pro Asp Leu Ser Asn Cys Ser Ser Pro Trp Val		
545	550	555	
Asn His Ile Thr	Gln Lys Leu Lys Ser Gly Glu Thr Ala Ala Asn		
560	565	570	
Ile Ala Arg Glu	Leu Ala Glu Gln Thr Arg Asn His Leu Asn Ala		
575	580	585	
Gly Asp Ile Thr	Tyr Ser Val Arg Ala Met Asp Gln Leu Val Gly		
590	595	600	
Leu Leu Asp Val	Gln Leu Arg Asn Leu Thr Pro Gly Gly Lys Asp		
605	610	615	
Ser Ala Ala Arg	Ser Leu Asn Lys Leu Gln Lys Arg Glu Arg Ser		
620	625	630	
Cys Arg Ala Tyr	Val Gln Ala Met Val Glu Thr Val Asn Asn Leu		
635	640	645	
Leu Gln Pro Gln	Ala Leu Asn Ala Trp Arg Asp Leu Thr Thr Ser		
650	655	660	
Asp Gln Leu Arg	Ala Ala Thr Met Leu Leu His Thr Val Glu Glu		
665	670	675	
Ser Ala Phe Val	Leu Ala Asp Asn Leu Leu Lys Thr Asp Ile Val		
680	685	690	
Arg Glu Asn Thr	Asp Asn Ile Lys Leu Glu Val Ala Arg Leu Ser		
695	700	705	
Thr Glu Gly Asn	Leu Glu Asp Leu Lys Phe Pro Glu Asn Met Gly		
710	715	720	
His Gly Ser Thr	Ile Gln Leu Ser Ala Asn Thr Leu Lys Gln Asn		
725	730	735	
Gly Arg Asn Gly	Glu Ile Arg Val Ala Phe Val Leu Tyr Asn Asn		
740	745	750	

WO 02/10387

PCT/US01/23433

Leu Gly Pro Tyr Leu Ser Thr Glu Asn Ala Ser Met Lys Leu Gly
 755 760 765
 Thr Glu Ala Leu Ser Thr Asn His Ser Val Ile Val Asn Ser Pro
 770 775 780
 Val Ile Thr Ala Ala Ile Asn Lys Glu Phe Ser Asn Lys Val Tyr
 785 790 795
 Leu Ala Asp Pro Val Val Phe Thr Val Lys His Ile Lys Gln Ser
 800 805 810
 Glu Glu Asn Phe Asn Pro Asn Cys Ser Phe Trp Ser Tyr Ser Lys
 815 820 825
 Arg Thr Met Thr Gly Tyr Trp Ser Thr Gln Gly Cys Arg Leu Leu
 830 835 840
 Thr Thr Asn Lys Thr His Thr Thr Cys Ser Cys Asn His Leu Thr
 845 850 855
 Asn Phe Ala Val Leu Met Ala His Val Glu Val Lys His Ser Asp
 860 865 870
 Ala Val His Asp Leu Leu Asp Val Ile Thr Trp Val Gly Ile
 875 880 885
 Leu Leu Ser Leu Val Cys Leu Leu Ile Cys Ile Phe Thr Phe Cys
 890 895 900
 Phe Phe Arg Gly Leu Gln Ser Asp Arg Asn Thr Ile His Lys Asn
 905 910 915
 Leu Cys Ile Ser Leu Phe Val Ala Glu Leu Leu Phe Leu Ile Gly
 920 925 930
 Ile Asn Arg Thr Asp Gln Pro Ile Ala Cys Ala Val Phe Ala Ala
 935 940 945
 Leu Leu His Phe Phe Phe Leu Ala Ala Phe Thr Trp Met Phe Leu
 950 955 960
 Glu Gly Val Gln Leu Tyr Ile Met Leu Val Glu Val Phe Glu Ser
 965 970 975
 Glu His Ser Arg Arg Lys Tyr Phe Tyr Leu Val Gly Tyr Gly Met
 980 985 990
 Pro Ala Leu Ile Val Ala Val Ser Ala Ala Val Asp Tyr Arg Ser
 995 1000 1005
 Tyr Gly Thr Asp Lys Val Cys Trp Leu Arg Leu Asp Thr Tyr Phe
 1010 1015 1020
 Ile Trp Ser Phe Ile Gly Pro Ala Thr Leu Ile Ile Met Leu Asn
 1025 1030 1035
 Val Ile Phe Leu Gly Ile Ala Leu Tyr Lys Met Val His His Thr
 1040 1045 1050
 Ala Ile Leu Lys Pro Glu Ser Gly Cys Leu Asp Asn Ile Asn Tyr
 1055 1060 1065
 Glu Asp Asn Arg Pro Phe Ile Lys Ser Trp Val Ile Gly Ala Ile
 1070 1075 1080
 Ala Leu Leu Cys Leu Leu Gly Leu Thr Trp Ala Phe Gly Leu Met
 1085 1090 1095
 Tyr Ile Asn Glu Ser Thr Val Ile Met Ala Tyr Leu Phe Thr Ile
 1100 1105 1110
 Phe Asn Ser Leu Gln Gly Met Phe Ile Phe Ile Phe His Cys Val
 1115 1120 1125
 Leu Gln Lys Lys Val Arg Lys Glu Tyr Gly Lys Cys Leu Arg Thr
 1130 1135 1140
 His Cys Cys Ser Gly Lys Ser Thr Glu Ser Ser Ile Gly Ser Gly
 1145 1150 1155
 Lys Thr Ser Gly Ser Arg Thr Pro Gly Arg Tyr Ser Thr Gly Ser
 1160 1165 1170
 Gln Ser Arg Ile Arg Arg Met Trp Asn Asp Thr Val Arg Lys Gln
 1175 1180 1185
 Ser Glu Ser Ser Phe Ile Thr Gly Asp Ile Asn Ser Ser Ala Ser
 1190 1195 1200
 Leu Asn Arg Glu Gly Leu Leu Asn Asn Ala Arg Asp Thr Ser Val
 1205 1210 1215
 Met Asp Thr Leu Pro Leu Asn Gly Asn His Gly Asn Ser Tyr Ser

WO 02/10387

PCT/US01/23433

1220 1225 1230
 Ile Ala Ser Gly Glu Tyr Leu Ser Asn Cys Val Gln Ile Ile Asp
 1235 1240 1245
 Arg Gly Tyr Asn His Asn Glu Thr Ala Leu Glu Lys Lys Ile Leu
 1250 1255 1260
 Lys Glu Leu Thr Ser Asn Tyr Ile Pro Ser Tyr Leu Asn Asn His
 1265 1270 1275
 Glu Arg Ser Ser Glu Gln Asn Arg Asn Leu Met Asn Lys Leu Val
 1280 1285 1290
 Asn Asn Leu Gly Ser Gly Arg Glu Asp Asp Ala Ile Val Leu Asp
 1295 1300 1305
 Asp Ala Thr Ser Phe Asn His Glu Glu Ser Leu Gly Leu Glu Leu
 1310 1315 1320
 Ile His Glu Glu Ser Asp Ala Pro Leu Leu Pro Pro Arg Val Tyr
 1325 1330 1335
 Ser Thr Glu Asn His Gln Pro His His Tyr Thr Arg Arg Arg Ile
 1340 1345 1350
 Pro Gln Asp His Ser Glu Ser Phe Phe Pro Leu Leu Thr Asn Glu
 1355 1360 1365
 His Thr Glu Asp Leu Gln Ser Pro His Arg Asp Ser Leu Tyr Thr
 1370 1375 1380
 Ser Met Pro Thr Leu Ala Gly Val Ala Ala Thr Glu Ser Val Thr
 1385 1390 1395
 Thr Ser Thr Gln Thr Glu Pro Pro Pro Ala Lys Cys Gly Asp Ala
 1400 1405 1410
 Glu Asp Val Tyr Tyr Lys Ser Met Pro Asn Leu Gly Ser Arg Asn
 1415 1420 1425
 His Val His Gln Leu His Thr Tyr Tyr Gln Leu Gly Arg Gly Ser
 1430 1435 1440
 Ser Asp Gly Phe Ile Val Pro Pro Asn Lys Asp Gly Thr Pro Pro
 1445 1450 1455
 Glu Gly Ser Ser Lys Gly Pro Ala His Leu Val Thr Ser Leu
 1460 1465

<210> 10
 <211> 469
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7477349CD1

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

WO 02/10387

PCT/US01/23433

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

<210> 11

<211> 335

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 55002225CD1

<400> 11

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

WO 02/10387

PCT/US01/23433

```

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

<210> 12
<211> 630
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475686CD1

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

```

WO 02/10387

PCT/US01/23433

	95		100		105
Gly Asp Phe Gln Val Tyr Gly Leu Val Ser Thr Glu Gly Val Val	110		115		120
Ser Cys Phe Gly Glu Lys Thr Pro Gln His Pro Gly Pro Pro Ala	125		130		135
Ser Leu Ser Leu Ala Asn Arg Cys His Asn Val Val Thr Ala Val	140		145		150
Gly Ala Trp Pro Ala His Gly Ser Ile Leu Gly Asn Val Pro Glu	155		160		165
Ala Pro Val Gly Ala Asp Val Leu Gly Ala Gly Gly Cys Asp Trp	170		175		180
Ala Asp Lys Glu Ala Leu Ala Pro Gly Gln Arg Ala Lys Val His	185		190		195
Ile Leu Leu Glu Ser Ser Gly Gln Ser Asp Pro Ser Tyr Ala Val	200		205		210
Leu Pro Asp Ser Trp Ala Ala Thr Glu Gly Phe Pro Thr Tyr Arg	215		220		225
Ser Gln Val Ser Ser Pro Arg Ile Pro Gly Ser Ser Ile Trp Leu	230		235		240
Gly Ser Gly Ser Gly Trp Pro Ile Leu Gly Glu Leu Arg Glu Cys	245		250		255
Asp Gln Met Phe Ser Cys Met Leu Pro Thr Gly Cys Ala Ser Phe	260		265		270
Gln Asp Pro Gly Arg Tyr Gly Asp Tyr Asp Leu Pro Met Asp Glu	275		280		285
Asp Glu Asp Met Thr Lys Thr Arg Thr Phe Phe Ala Ala Lys Ile	290		295		300
Val Ile Gly Ile Ala Leu Ala Gly Ile Met Leu Val Cys Gly Ile	305		310		315
Gly Asn Phe Val Phe Ile Ala Ala Leu Thr Arg Tyr Lys Lys Leu	320		325		330
Arg Asn Leu Thr Asn Leu Leu Ile Ala Asn Leu Ala Ile Ser Asp	335		340		345
Phe Leu Val Ala Ile Ile Cys Cys Pro Phe Glu Met Asp Tyr Tyr	350		355		360
Val Val Arg Gln Leu Ser Trp Glu His Gly His Val Leu Cys Ala	365		370		375
Ser Val Asn Tyr Leu Arg Thr Val Ser Leu Tyr Val Ser Thr Asn	380		385		390
Ala Leu Leu Ala Ile Ala Ile Asp Arg Tyr Leu Ala Ile Val His	395		400		405
Pro Leu Lys Pro Arg Met Asn Tyr Gln Thr Ala Ser Phe Leu Ile	410		415		420
Ala Leu Val Trp Met Val Ser Ile Leu Ile Ala Ile Pro Ser Ala	425		430		435
Tyr Phe Ala Thr Glu Thr Val Leu Phe Ile Val Lys Ser Gln Glu	440		445		450
Lys Ile Phe Cys Gly Gln Ile Trp Pro Val Asp Gln Gln Leu Tyr	455		460		465
Tyr Lys Ser Tyr Phe Leu Phe Ile Phe Gly Val Glu Phe Val Gly	470		475		480
Pro Val Val Thr Met Thr Leu Cys Tyr Ala Arg Ile Ser Arg Glu	485		490		495
Leu Trp Phe Lys Ala Val Pro Gly Phe Gln Thr Glu Gln Ile Arg	500		505		510
Lys Arg Leu Arg Cys Arg Arg Lys Thr Val Leu Val Leu Met Cys	515		520		525
Ile Leu Thr Ala Tyr Val Leu Cys Trp Ala Pro Phe Tyr Gly Phe	530		535		540
Thr Ile Val Arg Asp Phe Phe Pro Thr Val Phe Val Lys Glu Lys	545		550		555
His Tyr Leu Thr Ala Phe Tyr Val Val Glu Cys Ile Ala Met Ser	560		565		570

WO 02/10387

PCT/US01/23433

```

Asn Ser Met Ile Asn Thr Val Cys Phe Val Thr Val Lys Asn Asn
575 580
Thr Met Lys Tyr Phe Lys Lys Met Met Leu Leu His Trp Arg Pro
590 595
Ser Gln Arg Gly Ser Lys Ser Ser Ala Asp Leu Asp Leu Arg Thr
600 610
Asn Gly Val Pro Thr Thr Glu Glu Val Asp Cys Ile Arg Leu Lys
620 625

```

```

<210> 13
<211> 695
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7482007CD1

```

```

<400> 13
Met Lys Met Lys Ser Gln Ala Thr Met Ile Cys Cys Leu Val Phe
1 5 10 15
Phe Leu Ser Thr Glu Cys Ser His Tyr Arg Ser Lys Ile His Leu
20 25 30
Lys Ala Gly Asp Lys Leu Gln Ser Pro Glu Gly Lys Pro Lys Thr
35 40 45
Gly Arg Ile Gln Glu Lys Cys Glu Gly Pro Cys Ile Ser Ser Ser
50 55 60
Asn Cys Ser Gln Pro Cys Ala Lys Asp Phe His Gly Glu Ile Gly
65 70 75
Phe Thr Cys Asn Gln Lys Lys Trp Gln Lys Ser Ala Glu Thr Cys
80 85 90
Thr Ser Leu Ser Val Glu Lys Leu Phe Lys Asp Ser Thr Gly Ala
95 100 105
Ser Arg Leu Ser Val Ala Ala Pro Ser Ile Pro Leu His Ile Leu
110 115 120
Asp Phe Arg Ala Pro Glu Thr Ile Glu Ser Val Ala Gln Gly Ile
125 130 135
Arg Lys Asn Cys Pro Phe Asp Tyr Ala Cys Ile Thr Asp Met Val
140 145 150
Lys Ser Ser Glu Thr Thr Ser Gly Asn Ile Ala Phe Ile Val Glu
155 160 165
Leu Leu Lys Asn Ile Ser Thr Asp Leu Ser Asp Asn Val Thr Arg
170 175 180
Glu Lys Met Lys Ser Tyr Ser Glu Val Ala Asn His Ile Leu Asp
185 190 195
Thr Ala Ala Ile Ser Asn Trp Ala Phe Ile Pro Asn Lys Asn Ala
200 205 210
Ser Ser Asp Leu Leu Gln Ser Val Asn Leu Phe Ala Arg Gln Leu
215 220 225
His Ile His Asn Asn Ser Glu Asn Ile Val Asn Glu Leu Phe Ile
230 235 240
Gln Thr Lys Gly Phe His Ile Asn His Asn Thr Ser Glu Lys Ser
245 250 255
Leu Asn Phe Ser Met Ser Met Asn Asn Thr Thr Glu Asp Ile Leu
260 265 270
Gly Met Val Gln Ile Pro Arg Gln Glu Leu Arg Lys Leu Trp Pro
275 280 285
Asn Ala Ser Gln Ala Ile Ser Ile Ala Phe Pro Thr Leu Gly Ala
290 295 300
Ile Leu Arg Glu Ala His Leu Gln Asn Val Ser Leu Pro Arg Gln
305 310 315
Val Asn Gly Leu Val Leu Ser Val Val Leu Pro Glu Arg Leu Gln

```

WO 02/10387

PCT/US01/23433

320 325 330
 Glu Ile Ile Leu Thr Phe Glu Lys Ile Asn Lys Thr Arg Asn Ala
 335 340 345
 Arg Ala Gln Cys Val Gly Trp His Ser Lys Lys Arg Arg Trp Asp
 350 355 360
 Glu Lys Ala Cys Gln Met Met Leu Asp Ile Arg Asn Glu Val Lys
 365 370 375
 Cys Arg Cys Asn Tyr Thr Ser Val Val Met Ser Phe Ser Ile Leu
 380 385 390
 Met Ser Ser Lys Ser Met Thr Asp Lys Val Leu Asp Tyr Ile Thr
 395 400 405
 Cys Ile Gly Leu Ser Val Ser Ile Leu Ser Leu Val Leu Cys Leu
 410 415 420
 Ile Ile Glu Ala Thr Val Trp Ser Arg Val Val Val Thr Glu Ile
 425 430 435
 Ser Tyr Met Arg His Val Cys Ile Val Asn Ile Ala Val Ser Leu
 440 445 450
 Leu Thr Ala Asn Val Trp Phe Ile Ile Gly Ser His Phe Asn Ile
 455 460 465
 Lys Ala Gln Asp Tyr Asn Met Cys Val Ala Val Thr Phe Phe Ser
 470 475 480
 His Phe Phe Tyr Leu Ser Leu Phe Phe Trp Ile Leu Phe Lys Ala
 485 490 495
 Leu Leu Ile Ile Tyr Gly Ile Leu Val Ile Phe Arg Arg Met Met
 500 505 510
 Lys Ser Arg Met Met Val Ile Gly Phe Ala Ile Gly Tyr Gly Cys
 515 520 525
 Pro Leu Ile Ile Ala Val Thr Thr Val Ala Ile Thr Gly Pro Val
 530 535 540
 Lys Gly Tyr Met Arg Pro Glu Ala Cys Trp Leu Asn Trp Asp Asn
 545 550 555
 Thr Lys Ala Leu Leu Ala Phe Ala Ile Pro Ala Phe Val Ile Val
 560 565 570
 Ala Val Asn Leu Ile Val Val Leu Val Val Ala Val Asn Thr Gln
 575 580 585
 Arg Pro Ser Ile Gly Ser Ser Lys Ser Gln Asp Val Val Ile Ile
 590 595 600
 Met Arg Ile Ser Lys Asn Val Ala Ile Leu Thr Pro Leu Leu Gly
 605 610 615
 Leu Thr Trp Gly Phe Gly Ile Ala Thr Leu Ile Glu Gly Thr Ser
 620 625 630
 Leu Thr Phe His Ile Ile Phe Ala Leu Leu Asn Ala Phe Gln Gly
 635 640 645
 Phe Phe Ile Leu Leu Phe Gly Thr Ile Met Asp His Lys Ile Arg
 650 655 660
 Asp Ala Leu Arg Met Arg Met Ser Ser Leu Lys Gly Lys Ser Arg
 665 670 675
 Ala Ala Glu Asn Ala Ser Leu Gly Pro Thr Asn Gly Ser Lys Leu
 680 685 690
 Met Asn Arg Gln Gly
 695

<210> 14
 <211> 633
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 6769042CD1

<400> 14
 Met Tyr Phe Thr Ala Ala Ile Gly Lys His Ala Leu Leu Ser Ser

WO 02/10387

PCT/US01/23433

```

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

```

WO 02/10387

PCT/US01/23433

Phe Val Ile Val Val Asn Ile Gly Ile Leu Ile Ala Val Thr Arg
 485 490 495
 Val Ile Ser Gln Ile Ser Ala Asp Asn Tyr Lys Ile His Gly Asp
 500 505 510
 Pro Ser Ala Phe Lys Leu Thr Ala Lys Ala Val Ala Val Leu
 515 520 525
 Pro Ile Leu Gly Thr Ser Trp Val Phe Gly Val Leu Ala Val Asn
 530 535 540
 Gly Cys Ala Val Val Phe Gln Tyr Met Phe Ala Thr Leu Asn Ser
 545 550 555
 Leu Gln Gly Leu Phe Ile Phe Leu Phe His Cys Leu Leu Asn Ser
 560 565 570
 Glu Val Arg Ala Ala Phe Lys His Lys Ile Lys Val Trp Ser
 575 580 585
 Thr Ser Ser Ser Ala Arg Thr Ser Asn Ala Lys Pro Phe His
 590 595 600
 Asp Leu Met Asn Gly Thr Arg Pro Gly Met Ala Ser Thr Lys
 605 610 615
 Ser Pro Trp Asp Lys Ser Ser His Ser Ala His Arg Val Asp
 620 625 630
 Ser Ala Val

<210> 15
 <211> 370
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7476053CD1

<400> 15
 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 Asp Ala
 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 Ala Arg Val
 170 175 180
 Pro Ser Ala Leu Ile Ala Leu Ala Pro 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

WO 02/10387

PCT/US01/23433

Leu Gly Val Val Phe Val Tyr Arg Lys Ile Tyr Glu Ala Ala 240
 230
 Lys Phe Arg Phe Gly Arg Arg Arg Arg Ala Val Leu Pro Leu Pro 245
 245
 Ala Thr Met Gln Val Lys Glu Ala Pro Asp Glu Ala Glu Val Val 250
 255
 Phe Thr Ala His Cys Lys Ala Thr Val Ser Phe Gln Val Ser Gly 260
 265
 270
 275
 Asp Ser Trp Arg Glu Gln Lys Glu Arg Arg Ala Ala Met Met Val 280
 285
 290
 295
 Gly Ile Leu Ile Gly Val Phe Val Leu Cys Trp Ile Pro Phe Phe 300
 305
 310
 315
 Leu Thr Glu Leu Ile Ser Pro Leu Cys Ala Cys Ser Leu Pro Pro 320
 325
 330
 Ile Trp Lys Ser Ile Phe Leu Trp Leu Gly Tyr Ser Asn Ser Phe 335
 340
 345
 Phe Asn Pro Leu Ile Tyr Thr Ala Phe Asn Lys Asn Tyr Asn Asn 350
 355
 360
 Ala Phe Lys Ser Leu Phe Thr Lys Gln Arg 365
 370

<210> 16

<211> 324

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7480410CD1

<400> 16

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

WO 02/10387

PCT/US01/23433

Thr Cys Ser Ser His Ile Met Val Val Ser Val Phe Tyr Gly Ala
 245 250 255
 Ala Phe Tyr Thr Asn Val Leu Pro His Ser Tyr His Thr Pro Glu
 260 265 270
 Lys Asp Lys Val Val Ser Ala Phe Tyr Thr Ile Leu Thr Pro Met
 275 280 285
 Leu Asn Pro Leu Ile Tyr Ser Leu Arg Asn Lys Asp Val Ala Ala
 290 295 300
 Ala Leu Arg Lys Val Leu Gly Arg Cys Gly Ser Ser Gln Ser Ile
 305 310 315
 Arg Val Ala Thr Val Ile Arg Lys Gly
 320

<210> 17
 <211> 315
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 55036418CD1

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

WO 02/10387

PCT/US01/23433

Ala	Leu	Arg	Lys	Gly	Leu	Asp	Arg	Cys	Arg	Ile	Gly	Ser	Gln	His
				305						310				315

<210> 18
 <211> 324
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7481701CD1

<400> 18
 Met Glu Ser Pro Asn Gln Thr Thr Ile Gln Glu Phe Ile Phe Ser
 1 5 10 15
 Ala Phe Pro Tyr Ser Trp Val Lys Ser Val Val Cys Phe Val Pro
 20 25 30
 Leu Leu Phe Ile Tyr Ala Phe Ile Val Val Gly Asn Leu Val Ile
 35 40 45
 Ile Thr Val Val Gln Leu Asn Thr His Leu His Thr Pro Met Tyr
 50 55 60
 Thr Phe Ile Ser Ala Leu Ser Phe Leu Glu Ile Trp Tyr Thr Thr
 65 70 75
 Ala Thr Ile Pro Lys Met Leu Ser Ser Leu Leu Ser Glu Arg Ser
 80 85 90
 Ile Ser Phe Asn Gly Cys Leu Leu Gln Met Tyr Phe Phe His Ser
 95 100 105
 Thr Gly Ile Cys Glu Val Cys Leu Leu Thr Val Met Ala Phe Asp
 110 115 120
 His Tyr Leu Ala Ile Cys Ser Pro Leu His Tyr Pro Ser Ile Met
 125 130 135
 Thr Pro Lys Leu Cys Thr Gln Leu Thr Leu Ser Cys Cys Val Cys
 140 145 150
 Gly Phe Ile Thr Pro Val Pro Glu Ile Ala Trp Ile Ser Thr Leu
 155 160 165
 Pro Phe Cys Gly Ser Asn His Leu Glu His Ile Phe Cys Asp Phe
 170 175 180
 Leu Pro Val Leu Arg Leu Ala Cys Thr Asp Thr Arg Ala Ile Val
 185 190 195
 Met Ile Gln Val Val Asp Val Ile His Ala Val Glu Ile Ile Thr
 200 205 210
 Ala Val Met Leu Ile Phe Met Ser Tyr Asp Gly Ile Val Ala Val
 215 220 225
 Ile Leu Arg Ile His Ser Ala Gly Gly Arg Thr Ala Phe Ser
 230 235 240
 Thr Cys Val Ser His Phe Ile Val Phe Ser Leu Phe Phe Gly Ser
 245 250 255
 Val Thr Leu Met Tyr Leu Arg Phe Ser Ala Thr Tyr Ser Leu Phe
 260 265 270
 Trp Asp Ile Ala Ile Ala Leu Ala Phe Ala Val Leu Ser Pro Phe
 275 280 285
 Phe Asn Pro Ile Ile Tyr Ser Leu Arg Asn Lys Glu Ile Lys Glu
 290 295 300
 Ala Ile Lys Lys His Ile Gly Gln Ala Lys Ile Phe Phe Ser Val
 305 310 315
 Arg Pro Gly Thr Ser Ser Lys Ile Phe
 320

<210> 19
 <211> 312
 <212> PRT
 <213> Homo sapiens

WO 02/10387

PCT/US01/23433

<220>
<221> misc_feature
<223> Incyte ID No: 7481774CD1

```

<400> 19
Met Glu Pro Trp Gln His Pro Thr His Phe Ile Leu Leu Gly Phe
1      5      10      15
Ser Asp Arg Pro His Leu Glu Arg Ile Leu Phe Val Val Ile Leu
20     25     30
Ile Ala Tyr Leu Leu Thr Leu Val Gly Asn Thr Thr Ile Ile Leu
35     40     45
Val Ser Arg Leu Asp Pro His Leu His Thr Pro Met Tyr Phe Phe
50     55     60
Leu Ala His Leu Ser Phe Leu Asp Leu Ser Phe Thr Thr Ser Ser
65     70     75
Ile Pro Gln Leu Leu Tyr Asn Leu Asn Gly Cys Asp Lys Thr Ile
80     85     90
Ser Tyr Met Gly Cys Ala Ile Gln Leu Phe Leu Phe Leu Gly Leu
95     100    105
Gly Gly Val Glu Cys Leu Leu Leu Ala Val Met Ala Tyr Asp Arg
110    115    120
Cys Val Ala Ile Cys Lys Pro Leu His Tyr Met Val Ile Met Asn
125    130    135
Pro Arg Leu Cys Arg Gly Leu Val Ser Val Thr Trp Gly Cys Gly
140    145    150
Val Ala Asn Ser Leu Ala Met Ser Pro Val Thr Leu Arg Leu Pro
155    160    165
Arg Cys Gly His His Glu Val Asp His Phe Leu Cys Glu Met Pro
170    175    180
Ala Leu Ile Arg Met Ala Cys Ile Ser Thr Val Ala Ile Asp Gly
185    190    195
Thr Val Phe Val Leu Ala Val Gly Val Val Leu Ser Pro Leu Val
200    205    210
Phe Ile Leu Leu Ser Tyr Ser Tyr Ile Val Arg Ala Val Leu Gln
215    220    225
Ile Arg Ser Ala Ser Gly Arg Gln Lys Ala Phe Gly Thr Cys Gly
230    235    240
Ser His Leu Thr Val Val Ser Leu Phe Tyr Gly Asn Ile Ile Tyr
245    250    255
Met Tyr Met Gln Pro Gly Ala Ser Ser Ser Gln Asp Gln Gly Lys
260    265    270
Phe Leu Thr Leu Phe Tyr Asn Ile Val Thr Pro Leu Leu Asn Pro
275    280    285
Leu Ile Tyr Thr Leu Arg Asn Arg Glu Val Lys Gly Ala Leu Gly
290    295    300
Arg Leu Leu Leu Gly Lys Arg Glu Leu Gly Lys Glu
305    310

```

<210> 20
<211> 1076
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474806CB1

```

<400> 20
caaaagttgaa tgcccgggttg gggcagagcc tgnatgccgtg ctgaggtcat gatgttatgc 60
tgtccatttt gcttccttcc aggggaagca gaagcgggag ccgtcgtgga gctctgctcc 120
tggaggggagc ctcccgggac atggagaagg tggacatgaa tacatcacag gaacaaggtc 180
ctcgcagtt ctacagagaag tacaagcaag tctacctctc cctggcctac agtatcatct 240
ttatctcagg gctgccaacta aatggcactg tcttgtggca ctctgggggc caaaccaagc 300

```

WO 02/10387

PCT/US01/23433

```

gctggagctg tgcaccacc tatctgggta acctgatggt ggccgacctg ctttatgtgc 360
tattgcccct cctcactc acctactcac tagatgacag gtggcccttc ggggagctgc 420
tctgcaagct ggtgcacttc ctgttctata tcaaccttta cggcagcctc ctgctgctga 480
cctgcatctc tgtgaccacg ttccctaggtg tgtggcaccg actgtgtttg ctgcccacc 540
ggcccccag gcatgacctg ctgggaccg gcaaccactg ggccctagtg gctctcagc 600
tgetgcccac actggccttc tcccaacgg actacatcaa tggccagctg atctggtagt 660
acatgaccag ccaagagaat tttgatcggc tttttgccta cggcatagtt ctgacattgt 720
ctggcttttt ttcccctcc ttggctattt tgggtgtgta ttcactgatg gtcaggagcc 780
tgatcaagcc aggggagac ctcactgagga caggcaaac agcccagcc aggtccatcc 840
ggaccatcct actggtgtgt ggcctcttca cctctgttt tgtgcccttc catatcactc 900
gctcctteta cctcaccato tgccttctgc tttctcagga ctgccagctc ttgatggcac 960
ccagtgggc ctacaagata tggaggctc tggtagtgt gaggagctgc ctcaaccagc 1020
tcctgtactt tctttcaagg ggggcaaaa tagagttagt ctctccaga aactga 1076

```

```

<210> 21
<211> 1102
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7474840CB1

```

```

<400> 21
ggaggctgag gcaggagaat ggcatacc caggaggcag agcttgcagt gagatgagat 60
catgccactg tctccagcc tggcacaac agcgagactc tgtctcaaaa aaaaaaaaaa 120
acaaaaaaaa aaaccttttt tcccaageta coattggact tttagccaac acctttttcc 180
ttttctcaa catcttcaata ttccttcagg atcagaaatc gaagcccat gacctcatca 240
gctgtaattc ggccttcaat catgtagtga tgttctcac tgggtggat gcttggctc 300
cagatattgc tgaatcctg cacttggga atgagttcaa atttaagtc ttttcttaca 360
taaacagagt gaggatgggc ctatgtatct gtaacatctg tctctgagt ataccagc 420
ccaaccactc cagcccac accttctgtt tggcaagctc taacagaaa ttcacaata 480
acattatcat gtcacttttt tttcttttt ttttttggc catcaattg tcttcaatt 540
ataaactagt attcttact gtggcttctt ctaatgtgac ccagacagt ctacctaa 600
gcagcaaac tgttcacttt ctcccataga agtccctcat gagaaaagta ttttttact 660
tgacattatc cagggatgct ttcattatag gaattacact gcattcaatt gcacacatg 720
tgatccttgt gtcagagcat gagacgcaat ctcagcaact tcacagcact agcatctctc 780
caaaccttt ccagagaaa agggctgctc agaccatccc gctgttagtg agctactgtc 840
tggctcagtg ctgggtggac ctcatcatct catcttcttc aacctgtctg tggcgtgta 900
accagctctt ctgagtagt cagaaccttg tggcgagatg ctatgccact gttgttctac 960
tggacaatat cagctctgat aaaaatatag ttgacattct ccaaaatag caaagtgcta 1020
taaaacttba caaagtggc gatggaaaac atttctaaaa aatagtcttc tcctatagt 1080
caattgttca agtagccctg ga

```

```

<210> 22
<211> 2529
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7475092CB1

```

```

<400> 22
ggctccggg tttccggccg ttcactgagc gaaaagagaa agcaaatgc cctcaggagg 60
ctccagcccg ccggagccc tccacggccg cggggggcag cggaggcga ggcgcgctg 120
ctcagactc aggcgcgagc gtgcccggct ccgtgcagtt ggcgctgagc gtcctgcaag 180
ccctctctca cgcgcgctg ttcgctttg cctacctgca gctgtggcgc ctgtcctct 240
accgcagagc gcggtgagt taccagagcc tctgctctt cctctgtctc ctgtggcag 300
cgtcaggac caccctctc tccgcgctc tctgctcag cggctccctg cctctgctc 360
ggcgcggcgc tcaactgac ttcttcccc actggctgct ctactgctc cctctgctc 420
tccagttctc cagctctgt ctctcaaac tctacctggc ggaggttata tgtaaagtca 480
gatgtgccac tgaacttgac agacacaaaa ttctactgca tttggcttt ataatggcaa 540

```

WO 02/10387

PCT/US01/23433

```

gcctgctctt tttagtggg aacttgactt ggcgaatgct agttcatgga gatgtcccag 600
aaaaacagtt gaagtgactt ggtttgttc gagcattaat taatgatagc ctggtttatc 650
ttgtggccat ctctttagtg tgttacatat gcaaaattac aaaaatgtca ttagctaatg 720
tttaactcga atcaaaaggt atgtctctgt gccaagatgt cgtctggggc tctgtagtca 780
ttctctgta ctctccaga gcttgttata atttgggtgt ggtaccata tctcaggata 840
cattagaagc tcaattaat tatggctggg ataattttc agataagctc catgtagaag 900
acataagtg agaaagat atagtattg gaatggtcct ctctctggg gaacatgtgc 960
cagcatggtt ggtgtagct ttttccggg cacagagatt aaaccagaat ttggcaactg 1020
ctggcatgat aaatagtcac agttatagtt ccagagctta ctttttcgac aatccaagac 1080
gatgatgat tgatgatgac ctgccaagac tgggaagttc aagagaagga agttaccaca 1140
attcgcgaag tttgggctgg tatggcacc aactgggttg tggcagcagc agttaccag 1200
tcaactccca cctgaaatgga cctatgacag atactgtccc tttgctcttt actttagta 1260
atttagattt gaacaatcat catagcttat atgtgacacc acaaaactga cagcatcacc 1320
aagtcatgat tcttgagttg ttttccataa atgtgtatat tcaatgtgtt taaattccat 1380
ctacataaac attccattat ctgttgcaac tgaacaacaa atctggaagt gtggctgtgt 1440
ttggbaaat scacagctat tatltttgac ctcttccabg taaaatgagc taaaatgaa 1500
agtttggagt aggaagaag agagattaga tcttaaggca ctgtatggcc tccaaaaac 1560
ctgactttgg aacatcaaat gcattatgac acttttatct ttgtctgag tcaactgag 1620
ccccaaagtc atatgcaat gttccactg aaatactgta ttgtaccaca aactggaag 1680
caattttcct atgaaatca aagccgtat attcattggt atgctctata cagatctctt 1740
aataaaat ttatagtggt aacagtgca agagttaag cataaaaatg tatcattctt 1800
tataaaatc tactgaaat gtgtaacat tgaagcagc tcttttaagc atgattttaa 1860
aatagcaact gaaatccaat cttttaaacc aaatgatggt agtaacctc tagttatggc 1920
cagcagtggt ctttgggag ccacaataat tcaagagga aataatcca gtgaaaatg 1980
tgtgctatt ttgtagaga ttggtcagtt gattattttg tgaattgag atatatgat 2040
tagtttaagc atgattcttg aagaagcaaa tagtgacttt tgcataagga gattttgta 2100
gaaacttctt gggactaac aagtttagag atgcatttaa gaattattca caaaatgtgt 2160
aatttaaat taacaataa atatttttc aaaaagcatt gatttctctg aagctgata 2220
tagctggctc taactagtg atcagatgtg tctcaggtta tctgaaatca tgatacata 2280
ttcagtgaa ctcaagtgca atactttgta agacataaa ttcctatgat ttccattt 2340
ttatattcta tatatggaa aagccaantt aaattgaatt cagattaatt ccagcattag 2400
actaaatgag caaacttaag taatgtaca aactaggtaa gtataaaacc acaggttaac 2460
aatattggag tacttttaga attacattaa aactgtctta aatgtctcat cccaaatca 2520
aaaaaaaa 2529

```

```

<210> 23
<211> 1847
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7341260CB1

```

```

<400> 23
ggggaggcca ggctcccag ttgctgcagt ttggaatag tcaagttcca ccctcccaga 60
ggcggggcca gggctgagct ctgcccagct catttctcta tccctctgag aaccagacg 120
ggcagagcct gggtaggaga gcttggcccc gctgtcccca ctgggtggag acacacatgca 180
cttgtccac ttgtctctt cagccaggac accagacatg gtcocaaacc ctgcagggtc 240
ggctgcagca actcctgac actcaggaa gcccaggctg ggcaggcaat acctgtccc 300
aacagccatg catgcccgtt gccgtcccag gactcccctg tcccagagc caagatgacg 360
cccacagca ctggcagggt gccacgccc attcccagg gggctttggt gctcccctg 420
gcoctgcaaa gctctatcct caccggaaac ctgctcttag cctcaggcat cgcctgggac 480
cgccgctgc gcagcccacc tgtctgctgc ttcttctgga gctactgct ggtcgggtg 540
ctcaaggtt tggcattgca cacattgcca gggctgtgga accagatcg ccgggttac 600
tggctctgct tctctctcta ctgtgctccc aacttctcct tctctcctc gcttcccac 660
ctcttctggt tgcaggggga gcgtacatg gcagtcctga ggccaatcca gccctctggg 720
agcattcggt tggccctgct cctcacctg gctggtcccc tgccttttgc cagtctgccc 780
gctctgggtt ggaaccactg gaccctggt gcccaactgca gctcccagcc tatcttccca 840
gccctctacc tgcctctgga agtctatggg ctctctgctc ccgctggtg tgcctgctcc 900
ttcctctctg tccgctgctt ggcactgccc caccgcccag tgcaggacat ctgcccgtg 960
gagcgggagc tgcctgcca tgcagcctcc gccctggccc gggcccttac ctggggcag 1020
gcaaggccac aggtcgggac catgctgctc ttggggctgt gctgggggccc ctacgtggcc 1080

```

WO 02/10387

PCT/US01/23433

```

acaactgctcc tctcagtcct ggcctatgag cagcgcggcc cactggggccc tgggacactg 1140
ttgtccctcc tctccctagg aagtgccagt gcagcggcag tgcccgtagc catggggctg 1200
gggatacaagc gctacacagc cccctggagg gcagccggccc aaaggtgccc gcaggggctg 1260
tggggaagag cctccgggga cagtcocggc cccagcattg cctaccaccg aagcagccaa 1320
agcagtgctcg acctggactt gaactaaagg aagggcctct gctgaetctc accagagcat 1380
ccgtccagctc cagccatcca gctgtctctc actgggcccc acttctctgg atcagagacc 1440
ctgcctctgt ttgacccccc actgaactgaa taagctctct ctggccgtta aaaaaaaaaa 1500
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaacaacag 1560
aaaaaaaaaa aaagaacac acaagaaca cagaacaac caagcagca accacacaca 1620
aaacaatgc acacacagaa caagacaca tcaagacag agagcacaca gcacggacc 1680
cagccacgcc cccagcactg accaccagc cccgacacag aaacyaacac tgaagactca 1740
acgcaaaaaa cgcacaaccg accacaagc aaccgctca cggcccagca acgaacacac 1800
atcaaaaaac aaaccgagac aaccacata cagccaaaaa aaccaca 1847

```

```

<210> 24
<211> 2031
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7473911C81

```

```

<400> 24
atgaataaaa acaacaaacc ttccagtttc atagccataa gaaatgctgc tttctctgaa 60
gtcggcattg ggatctctgc caatgccatg ctccctctct tccacatctc cactggcctt 120
ctcaagcaaca ggaccaagcc cgtgacctgc atcgtttctc atgtggctct aatccatato 180
atattgctgc taaccaacaga gttcatagct acagatattt ttgggtctca ggaltcagag 240
gatgacatca aacataagtc agttatctac aggggtaaca ggcagtccca gcattttcac 300
agcaacaacc ttctccaaa agcaccccca gaaaaatgg ccaagcagac catctctctg 360
ctcgtgagtt gctttgtgat tgtgatgttt ttggactgtg ttgtcgctc ctgctcagga 420
ctggtgtgga acagtgtacc agtccctcat cagctccaga tgctggtgga caatggctat 480
gccaccatca gtccctcagt gctaccaggc ctgactgccc caaacagtg gagagccagt 540
gtgtacctga atgacagctt gaacaaatgc agcaaggacc ggtctctctg tctagacagg 600
gggcttgatg aggggcccc gtccgtccca aagtgtctg agtcagagac cgcagagga 660
tacatcgtcc tcagggctcc gctgagggag gacgaaccca aggacggggc cagtgtgggg 720
aatgcagccc tgggtctccc cgaggcctct gcagaagagg aagagggagc tgaggagga 780
ggcagagcat gtggcctgga gaggacagga gctggtgggg agcaggttga ccttggtaga 840
ctactgacc atgaggagaa aagcaaccag aaagtggcag ctgccacctt ggaggaccgc 900
acacaggatg agcctgctga ggagagctgc cagatcgtcc ttttccagaa caactgcatg 960
gacaacttgy tgacttcctc cacaggaagc cctcagagt tcttcccaac caagagcacc 1020
tctttttgca gggagagctg ttctcctttt tctgtgtcag tgaaaagctt agaalcagag 1080
cagggaccaa agttggggct gtgtggggag gaggaccctc tggttggggc ttgtgtggc 1140
cagcatggac ccttgcaaga tggagtggcg gagggtccca cagcccttga tgtggtgtc 1200
ctgccgaagg agggagagaa ggaggagctc attgtggatg acatgctggc caaccctat 1260
gtgatgggag atgaggggga ggaggaggag gaggagttcg tggatgacac actggccaac 1320
ccctatgtag tgggagctgg cctgccagga agaggagggg aggagggga ggaggaggag 1380
ctcgtggatg acacgctggc cagcctctat aagatgggag aagaacatcg acacaagggc 1440
ctggcccacc tctgggaagg tggccagaaa ccgtcccaga aactgcccc aaagaacca 1500
gatctgagcc aggttctcca gcccttggca tcggaggtgc cgcagaggag gcaggaaaa 1560
gctgtgttca ctgaagggag gcccttggaa gccagcaggc ccttgcacg aaagcccagg 1620
gccttcaact tataccctcg gtcgttctcc gtggaaggcc aagagattcc tgtttccatc 1680
tctgttactt gggagccaga aggttcgggg ttgatgacc acagataaaa gaggaaagag 1740
gaactctctc ctgtgtgttc tggagtttc tcccagagaa accacttcc atccagggc 1800
acctccacgc ctctctcat ggtcgcaatc ccacctctt tgcactggc ctgcatcac 1860
aagaagccca tcaacaagag ctctccctct ctctgatgc acagcagctc cccgacaag 1920
tacaagaaga agaagctatc ctlttaagcg ttccctggcg tgatgttaa caagatggag 1980
agccacggca cgatgctca tgcctgtcat ccagcactt tgggaagctg a 2031

```

```

<210> 25
<211> 1130
<212> DNA
<213> Homo sapiens

```

WO 02/10387

PCT/US01/23433

<220>
<221> misc_feature
<223> Incyte ID No: 7474767CB1

<400> 25
ggggggcgcct catggagcac acgcaagccc acctcgcgc caacagctcg ctgtcttggg 60
ggccccccgg ctggcctgc ggcttgggtt tctgcccgt ggtctactac agcctcttgc 120
tggcctcgg ttaccagca aatatcttga cagtgaatc octtcccag ctgggggcaa 180
gaagcagaaa gtctcctac aactatctct tggcactgc tgcctccgac atctgtgtcc 240
tcttttcaat agtgtttgtg gacttctctg tggaaattt catcttgaac atgcagatgc 300
ctcaggtccc cgacaagatc atagaagtgc tggaaattc atccatccac acctccatat 360
ggattactyt accgttaacc attgacaggt atatcgtct ctgccaccgc ctcaagtacc 420
acacggctct ataccagcc cgcaccggga aagtcattgt aagtgtttac atcaactgct 480
tctcgaccag catcccctat tactggtggc ccaacatctg gactgaagac tacatcagca 540
cctctgtgca tcaactctc atctggatcc actgcttcc cgtctacctg gtgcctgct 600
ccatctctt catcttgaac tcaatcatg tgtacaagct caggaggaag agcaatttcc 660
gtctccgtgg ctactccagc gggaaagaca cgcacatctt gttccaccat acctccatct 720
ttgcccaact ttggccccc cgcatacaca tgattcttta ccactctat gggggcgcca 780
tcagaaacgc ctggctggta cacatcatgt cgcacatgce caacatgcta gccctctgta 840
acacagccat caactcttc ctctactgct tcatcagcaa cgggttccgc accaaggcag 900
ccgccacgct caaggcttcc ttcaagtccc agaagcaacc tgtacagttc tacaccaatc 960
ataacttttc cataacaagt agcccctgga tctgcgggc aaactcaac tgcatcaaga 1020
tgctggtgta ccagtatgac aaaaatggaa aacctataaa aagtcgtaat gacagcaaaa 1080
gctccctacca gtttgaagat gccattggag cttgtgtcat catcctgtga 1130

<210> 26
<211> 1202
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475815CB1

<400> 26
caggttctgc aatacaattg gaaacaactc attgctcctg cctatgaaga agtagaacta 60
tggtaaaaat aagaaaaatgc cttccaata atagggagtt gtactatgga agtaaatagg 120
aggcccagca atgggaatca tagtccagcc atgatggcat agtcatgga gaagagagag 180
gggtgggatg gcttccagga ggggtggaag aggggggtcca gtaactgaag gggaaaagat 240
gcatacagaat taattaatgt attttgatga tggcaatagt gttggttggg attggtgag 300
gtagaaataa ttgtgatatt ttgttgcct ttctccctag acatbaacta tgtgcttatt 360
ttcccataa gatgaataa acaacaaac ctctccagtt catagcata agaactgctg 420
ctttctctga agtcggcatt gggatctctg ccaatgcat gctcctctc ttccacatcc 480
tcaactgctc tctcaagcac aggaccaage cgcctgact gatcgtttgt catgtggctc 540
taataccatc catattgctg ctaccacagc agttcatagc tacagatatt ttgggtctc 600
agatltcaga ggaatgacac aaacataagt cagttatcta caggtaacag ttgatgagag 660
gcctctccat ttccaccacc tgcctgctga gtatctccc ggccatcacc tgacagccca 720
gaagctcctg ttggcagtg ttcaagatt ctacatcac caaccagtt gcttctctt 780
cgtcttcca catatccatt agtgacagct tcttagtctc cactcttccc atcaaaaatc 840
tggcctcaaa tagccttaca ttgtcaactc aatcctgctc tgctgggac ggctcaagcc 900
ccccctccag tggatacatg gtgattctct tgtccaggcg taacaggcag tcccagcatt 960
ttcacagcac caactttct ccaaaagcac cccagaaaa aatggccacg cagaccatc 1020
ttctgctcgt gaggttgctt gtgattggt atgttttggc ctgttctgct gccctctgct 1080
caggactggt gtgaaacgt gatccagttc gtcactgagt ccagatgctg gtggcaatg 1140
gctatgcaac catcagctct tcagtgtctg tcagtactga aaaaatgagt atcaaatgct 1200
ga 1202

<210> 27
<211> 2079
<212> DNA
<213> Homo sapiens

<220>

WO 02/10387

PCT/US01/23433

<221> misc_feature
<223> Incyte ID No: 60263275CB1

```

<400> 27
taccgacacag tagagagcctt ccagggtctg ctggcgtggg ataccogtac cacagaaatg 60
cagggaacat tgcctctctc aggcctctgc tttctgtgca gcctctttgg agctgtgact 120
cagaaacca aaacatttaa tgaatgtaca ccacctata gtgtatattg tggatbtaac 180
gctgtgtgtt acaatgtcga aggaagtttc tactgtcaat gtgtcccagg ataatagactg 240
cattctggga atgaacaatt cagtaattcc aatgagaaca ectgtcagga caccacctcc 300
tcaagacaa cccagggcag gaagagctcg caaagattg tggacaaatt tgaagtaact 360
ctcaccatc agactttatg gagaacagaa gggagacaag aaatctcatc cacagctacc 420
actattctcc gggatgttga atcgaaagtt ctgagaacct ccttgaaga tccagaaaca 480
aaagtctcga aaatccaaaa cgatagtgtc gctattgaaa ctcaagcgat tacagacaat 540
tgctctgaag aaagaagac atccaactg aacgtccaaa tgaactcaat ggacatccgt 600
tgcagtgaca tcatccaggg agacacacaa ggtcccagtg ccattgcctt tatctcatat 660
tcttctcttg gaaacatcat aaatgcaact ttttttgaag agatggataa gaaagatcaa 720
gtgtatctga actctcaggt tgtgagtcct gctattggag ccaaaagaaa cgtgtctctc 780
tccaagcttg tgaecgtgac ttccagcacc gtgaagatga cccocagtae caaaaaggtc 840
ttctgtgtct actggaagag cacagggcag ggcagccagt ggtccagaga tggctgcttc 900
ctgatacaag tgaacaagag tcaacccatg tgtaatgtca gtaoctgtc cagctctogt 960
gctctgatgg cctgaccag ccaggggag gatccctgac tgactgtcat caoctaogtg 1020
ggcttgaggt tctctctgct gtgctctctc ctggggcgcc tcacttttct cctgtgtaaa 1080
ggcattccaga acaccagcac ctcaactgat ctgcagctct cgctctgctt ctctctggcc 1140
cacctctctc tctctgtggg gattgatcga actgaaccca aggtgtctgt ctccatctc 1200
ggcgtgctct tgcactatct ctacctgccc gcttccacct ggatgctgct ggagggtgtg 1260
cacctctctc tcaactgacg gaacctgaca gtggtcaact actcaagcat caatagactc 1320
atgaagtgga tcaatgtccc agtcggctat ggcgttcccg ctgtgactgt ggccattctt 1380
gcagctctct ggcctcactc ttatggaact gctgactgat gctggctcca cctggaccag 1440
ggatctcatg ggaattctct tggccagtc tgtgcaattc tctctgcaaa tttagtattg 1500
tttactcttg tcttttggat ttgaaaaga aaacttctct cctctatag tgaagtgtca 1560
accatccaga acacaaggat gctgctcttc aaagcaacag ctcagetott catctctggc 1620
tgaactggtg gtctggctct gctacaggtg ggtccagctg cccaggtcat ggccactctc 1680
ttcaacata tcaacagcct ccaaggcttc ttoactctt tggctctact cctcctcagc 1740
cagcaggtcc agaacaata tcaaaagtgt tttagagaga tctgtaaaatc aaaaactctg 1800
tctgagacat acacacttcc cagcaagatg ggtcctgact caaaacccag tgagggggat 1860
gtttttccag gacaagtgnaa gaaaaatat taaaactaga atattcaact ccatatggaa 1920
aatcatatcc atggatctct ttggcattat gaagaatgaa gctaaggaaa agggaaattc 1980
ttaacatata catccttggg gagggaatgaa tcaaccttca ctcccaaac tgtttgttct 2040
ccacaatagg tctcaacaaa tgtgtggtaa attgcatta 2079

```

<210> 28
<211> 5324
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 60203310CB1

```

<400> 28
ggcggcagca gcagcagcaa gaggaggaga agcagcccca gctatgactg ccgcatytta 60
atagctcccg ctgggtccct cggctgctgc tggagacaga gcctgactcc gaagtctgtc 120
aactgtggac tggagagac atttgaacct tctttctctt tgcctccctc ttgcccct 180
tgggtgtgtg gaagcagga agttaaagga aggcgaacat ttgctctct tttctctcc 240
cctttctcgg tggctgtgta ggggaagaaa gggagagag actttttgtt gttgttctc 300
tgactggggt ctccacctc ctgctgcttt ctctgocctt cgattctcgt tatttgccc 360
gtatgtttgg gogtgtctgc acagggccc gcccgtcttt tgcccgggc tcaatggctg 420
gatttggaaa actgocccc ccttcaggtt gttgagcaac tgatgggagc atctcaggga 480
ccggcgttta cgaagggttt cagatttggg atattgtgtt tctgttttg gagaattat 540
tcttttctct ttaatttga aaaaaatca tcaatgttgg aatacagaag agaaaataga 600
aatatagca ttttgtttca catttgaaca gtcatcttgg aggaatactc catacctgag 660
tagacagcca tgtggccctc gcagctacta attttcatga tgccttagc tcaataaatt 720
catgcttcca gccgtgcccc aattccaatg gctgtgtctc gcagagagct atcctgtgag 780

```

WO 02/10387

PCT/US01/23433

```

agctatccta tagagcttgc ctgtccagga acagacgtca tcatgataga aagtgccaac 840
tatggcagga ctgatgacaa aatttgtgac tctgaccctg ctcatagga gaataccga 900
tgttatctgc cagatgecta taagattatg tctcaaaagt gcaataacag aaccagctgt 960
gagtggtggy caggtctcga tgttttcca gaaccctgtc ccaggaccct taataccctt 1020
gaagtgcagt atgaatggt cctttacaaa gtggaacaaa aagltttct ttgtctgga 1080
ctactaaaag gegtatacca gagtgaacat ttgtttgagt ccgaccacca atctgggggg 1140
tgggtcaaaag accctctgca ggcactgac aagatttatt atatgcoctg gactccctac 1200
agaactgata ccoctgactga gtattatccc aaggatgact tcatgtctgg aagaaccaat 1260
acaacctaca agctccctca taggttggat ggcacaggat ttgtagtgtg tgatggagct 1320
ttgttcttca acaaaagagc caccaggaac atagtaaagt ttgatttgcg gactaggata 1380
aagagtggag aggcctatcat agcaaatgcc aattaccatg atacctcccc ttaccgatgg 1440
ggagggcaat ctgacataga cctggcagta gatgagaatg ggctatgggt aatctatgca 1500
acagaacaaa acaatggtaa aattgtcaat agtcaatgta acccttaaac cctaccggac 1560
gaaggaacat gggatctcgc atatgataaa aggtcagcctt ccaalgcctt tatgatttgt 1620
ggaaattctgt atgtggtcaa atctgtatat gaggatgatg acaatgaggg tactggaaat 1680
aagattgact acatttacia cactgaccaa agcaaggata gtttgggga tgtacccttt 1740
cctcaatlcat accagtaact tgcagctgtg aatttacaac ccagggacaa cctactttat 1800
gtatggata actatcaagt cgtgaataat tctttggatt ttggacctct ggtatgtaga 1860
tcaggggcag cscactatgg acaagtttca tacatttctc cggcaatcca ccttgactct 1920
ggcttagaaa gaccotctgt taaagatata tctaccaag gactctctgg catgggaagc 1980
actaccacca gtaccacctc teggaccaca accttggacc caggaaaggag taccaccctg 2040
tcagtgtcag gaagaagaaa cggagctact agtaccacct cctcagctgt cgaggtactt 2100
gatgacatga ccacacacct tccatcagca tcttcccaaa tcccagctct cgaagagagc 2160
tgtgaggtcg tggaaagccc agaaatcag ttgtttaaga ctctgcaagg acagatagca 2220
aagcagccat gccctcaggg aactataggt gtatcaacct atctatgctt tgcctctgat 2280
ggaaattggg atccccaaag tccagatctc agcaactggt cttctccttg ggtcaatcat 2340
ataacaacga agttgaaatc ttgtgaaaca gctgccaaca ttgttagaga gctggctgaa 2400
cagacaagaaa atcacttgaa tgcctggggac atcacctact ctgtcggggc catggaccag 2460
ctggtagccc tcttagatgt acagcttcgg aacttgaacc caggtggaaa agtatgtgct 2520
gcccggagtt tgacaagctc tcagaaaga gaggctctct gcagagcta tgtccagca 2580
atggtcagga cagttaacaa cctccttcag ccacaagctt tgaatgcatg gagagactg 2640
actacagagt atcagctcgc tgcggccacc atgttctctc atactgtgga ggaaggtgct 2700
tttgtgctgg ctgataacct tttgaagact gacattgtca gggagaacac agacaatatt 2760
aaatggag ttgcaagact gagcaagaaa ggaacttag aagacctaaa atttccagaa 2820
aacatggccc atggaaagcac tatccagctg tctgcaataa ccttaaagca aaatggccga 2880
aatggagaga tcagagtggc ctttgcctg tataacaact tgggtcctta tttatccagc 2940
gagaatgcca gtatgaagtt gggaaaggaa gctttgtcca caaatcatc ttgtattgtc 3000
aatccccctc ttattacggc agcaataaac aaagagtcca gtaacaaggt ttatttggct 3060
gatccctgtg tatttactgt taacaatata aagcagtcag aggaaaatit caacctaac 3120
tgttcaattt ggagctactc caagcttaca atgacaggtt atgtgtcaac acaaggtgt 3180
cggctcctga caacaataaa gacacatact acatgctctt gtaaccacct aacaaatctt 3240
gcaatctcga tggcaactgt ggaagttaag caagtgatg cggccaatga cctcctctg 3300
gatggatca cgtgggttgg aatttggctg tccctgtttt gctcctgat ttgcatctc 3360
acatttctct tttccggcg gctccagagt gaccgtaaca ccatccaca gaacctctg 3420
atcagttctc ttgtagcaga gctgctcttc ctgattggga tcaaccgaac tgaccaaccg 3480
attgctctg ctgtttctgc tgcctgttta cattctctc tcttggctgc cttcacctgg 3540
atgttctctg aggggtgca gctttatata atgctggtgg aggtttttga gagtgaacat 3600
tcagctagga aatacttcta totgctcggc tatgggatgc ctgcaactca ttgtgctgtg 3660
tcagctcag tagactacag gatttatgga acagataaag tatgttggct ccgacttgac 3720
acctacttca tttggagttt tataggacca gcaactttga taattatgct taatgtaac 3780
ttctctggga ttgctttata taaaatggtt catcactctg ctatactgaa acctgaatac 3840
ggctgtcttg ataacatcaa ctatgaggat aacagacctc tcatcaagtc atgggttata 3900
gggtgaaatg cttctctctg cctattagga ttgaactggg cctttggact catgtatatt 3960
aatgaaagca cagtaactab ggcctatctc ttaccattt tcaattctct acaggaatg 4020
tttatatta tttccattg tgcctacaq aagaaggtac gaaagagta tggaaatgc 4080
ctcggaaac attgtctgag tggcaaaagt acagaggtt ccattgttc agggaaacaa 4140
tctggttctc gaactcctg agctactcc acagctcac agagccgaat ccgtagaaatg 4200
tggaaatgaca cggctcgaag cagctcagag tcttcttata ttactggaga caLaaacagt 4260
tcagagctcac tcaacagaga ggggttctg aacaatgcca gggatacaag tytcatggat 4320
actctaccac tgaatggtaa ccatggcaat agttacagca ttgocagcg cgaatcctg 4380
agcaactctg tgcacaatcat agaccgtgac tataaccata acgagaccgc cctagagaaa 4440
aagattctga aggaactcac ttccaactat atcccttctt acctgaacaa ccatgagcgc 4500
tccagtgaac agaaacggaaa tctgatgaac aagctggtga ataaccttgg cagtggaaag 4560

```

WO 02/10387

PCT/US01/23433

```

gaagatgatg ccattgtcct ggatgatgcc acctcgttta accacgagga gagtttgggc 4620
ctggaaactca ttcatgagga atctgatgct cctttgtctgc cccaagagt atactccacc 4680
gagaaaccacc agccacacca ttataccaga agggggatccc cccaagacca cagttagagc 4740
tttttccctt tgcataccaa cgagcaacaca gaagatctcc agtcaaccca tagagactct 4800
ctctatacca gcatcccgac actggctggt g'tggccggca cagagaggtg taaccaacgc 4860
accagagcgc aacccccacc ggccaatgt ggtgatggcc aagatgttta ctacaaaagc 4920
atgccaaacc taggctccag aaaccaagtc catcagctgc atacttacta ccagctaggt 4980
cgggcgagca gtagtgagtt talagttcct ccaacaaaag atgggaacccc tcccgaggga 5040
agttcaaaag gaccggctca ttgttccact agtctataga agatgacaca gaaattggaa 5100
ccaacaaaac tgctaaaccc ttgttgactg ttctgagttg atataagcag tggtaataat 5160
gtgtgtactc ctaaatcttt atgctgtcct ctaaaagaca acacaaactc tcagactttt 5220
ttttttcaac tgggatttaa ggtcagccca ggggagaaag ataactgcta aaattccctc 5280
gtaccocact cttttctgtc cttttcccct tcagatggag actt 5324

```

```

<210> 29
<211> 1962
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7477349CBI

```

```

<400> 29
atggatccca gcggtgttag caatgagtat tatgatgttg cccatggagc aaaagatcca 60
gtggtcccca cttccctgca ggacatcaact gctgtcctgg gtaacagaagc atatactgag 120
gaagacaaat caatggtgtc ccatgcaacag aaaagccagc attctgtctc cagccatccc 180
aggtggctga ggtctccaca ggtcacaggg ggaagctggg acctccgaat aaggccatcc 240
aaggactcca gcagttcccg ccaggctcag tgtctgcgta aggatctagg ggcacaaac 300
cacttgaga gcccaagggt gagaggtaca gctggcgatg ctgacagggg gctgggggga 360
ccctcagaaa aggccacagc tggccagcca cagagtaccc tgbtgcacc ccccaactc 420
agcggctgga gccagagatt tgaagccac tggccagaga tgcagagag gtcctccgtg 480
gtggctggcg tcacccctgt catctactac agtgtcctgc tgggcttggg gctgctctgc 540
agcctcctga ccgcaagtgc cctggcgccg cttgccacca ggaccagagg gcctccctac 600
tactaacctc tggcgctcac agcctcggat atcatcatcc aggtggctca cgtgttcggg 660
ggcttccctc tgcagggagc agtgcctggc cgccaggctg cccaggctgt ggtgcgcacg 720
gccacaatcc tggagtttgc tgccaaccac gccctcagctt ggatgcacct cctgctcagc 780
gttgaccgct acactgccct gtgccacccc ctgcaccatc gggccgcctc gtcctccaggc 840
cggaccgcgc gggccattgc tgcctgctcg agtgcctgcc tgttgaocgg catccccctc 900
tactgggtgg tggacatgtg gagagacacc gactcaaccca gaacactgga cagagtcctc 960
aagtgggtct actgtctcac tctctatctc atcccctgtg gcctgttccc ggtccccaac 1020
tcggccatca tccaccgctc agggagaggg ggcggagtg ggttgcagcc ccgggtaggc 1080
aagagccacag ccactcctct ggcatcacc acactgttca cctcctgtg ggcgcccgg 1140
gtcttctgca tgcctacca catgtactgt gccctgtcc accgggactg gaggtccc 1200
ctggccttgg atgtggccaa catggtggcc atgctccaca cggcagccaa cttcgccctc 1260
tactgctttg tcagcaagac tttccgggcc actgtccgac aggtcatcca cgtgacctac 1320
ctgccctgca ctttggcctc acagccagag ggcattggcg cgaagcctgt gatggagcct 1380
ccgggaactc ccacaggggc agaagtgtag aggggggggc ccaagctagg agctcaggtt 1440
ggctcatggc cacatgtact gggcctcttg aggttgtacc caaaaacgtt ttatcaacag 1500
cttgccttcc ttgggtgggg gtaggagctc ctcccttggg tgtggctccc aggtagagag 1560
gaggacaact tagccagctc ttatgtttgc ttcaccagca atccctatit cctgggaaga 1620
tgaaaaggca ctgccaggca caggctaata gcactcagtc tgtgggcatt cctttgcggg 1680
gggcattttg cctggctcat cgtgaatgcc agatlaatgt tggctgaatg gatagaaaa 1740
cggctctcca ttttctaac tgaagcagga gaatcctgtt aaaccaaggg accgaggttg 1800
cagcagctg agatccggcc atagaaacac catggaactc caacctgggc aaccaaggtg 1860
aaactctgac tcaaaaaaaa aagagaaaaa aoncattagg taacagtttc tttttagcat 1920
ttgtgtaacc ttaataaaaa taaagtata atcaaaaaaa aa 1962

```

```

<210> 30
<211> 1558
<212> DNA
<213> Homo sapiens

```

WO 02/10387

PCT/US01/23433

<220>
<221> misc_feature
<223> Incyte ID No: 55002225CB1

<400> 30
atlttatttg cgaaggcacc ccagctcctc agaaaagagc acgacgcacc cgatgctcgg 60
attggatgaa gtggcaaacg tttaatccct gaaaagtcca cgaacsatga atccatttca 120
tgcattctgt tggcaacctc ctgcccgaact tttaaacaaa tccctggaata aagagtttgc 180
ttatacaact gccagtggtg tagalacagt catctccctc tccatgattg ggattatctg 240
ttcaacaggc ctggttggca acatccctcat tgtattcact ataataagat ccaggaaaaa 300
aacagtcctc gacatctata tetgcaacct ggctgtggct gatttggtcc acatagtctg 360
aatgctcttt ctatttcaac aatgggcccg agggggagag tgggtgtttg gggggcctct 420
ctgcaccatc atcacatccc tggatacttg taaccaatth gcccttagtg ccatcatgac 480
tgtaatgagt gtggacaggt actttgccct cgtccaacca tttcgactga ccagttggag 540
aacaaagtac aagaccatcc ggatcaatth gggcctttgg gcagcttctc ttatctggc 600
attgctgtc tgggtctact cgaaggtcat caaattttaa gacggtgttg agagtctgc 660
ttttgatctg acatccctcg acgatgtact ctggatata ctttatttga cgataacaac 720
tttttttttc cctctacccct tgaatttggg tgcctatctc ttaattttat gctatacttg 780
ggagatgatc caacagaata aggatgcccag atgctgcaat cccagtgtae caaacacagg 840
agtgatgaag ttgcaeaaga tgggtctggt gctgtgggta gtctttatcc tggatgctgc 900
cccttatcat gtgatcaaac tggatgaact acagatggaa cagcccacac tggccttcta 960
tgtgggttat tacccttcca tctgtctcag ctatgccagc agcagcatta acccttttct 1020
ctacatcctg ctgagtgaaa gccctcaaat ccaagaaga gcgactgaga aggaatcaa 1080
caatatggga aacactctga aatcacactt ttaggaaagt acatggatca ccatgagtct 1140
agacatgatt gtctatctta ctggtattat tagaaaaggg aggtgtaccg atatgtttat 1200
gccattctct ctgtgtact tgtgactctt agcagcatgg aagagaagtg taaccatgca 1260
aatcaaatga gcttaatatg ctaactttag caagatgtaa aatgttgatc tatattgtgg 1320
gtagggaatg ggatagtctg agataccocg gcttcatgat ggtgtatatt atttcagcat 1380
attataaaact agtcaactat gaaaatggcc atccatgacc attgactcaa aactcaccaa 1440
ggaaactgac ctggcctccc acactgcggc ctccactgtaa cagtttctct aaggttctca 1500
ggagggtatc accttagagt gaagtctaaa atttggctat tttttatota ttaaaaat 1558

<210> 31
<211> 2304
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475686CB1

<400> 31
atcggtgtgg gacctgtccc agcccggggc cgcgccctct tgtcttgggt tagggggctg 60
gaaagccgag gaggggagtg gaccaaatgc attgttcagc tgggtatctc ccttgcacc 120
cagcaccceq cggcggccac atgtggaatc gtttccagcg ccttggctat gcactcaaca 180
gatgtctgtc tggcccacac tatgcaaccg gcactggact gggcagcagc aatttggttt 240
acgggaagat taggactcac agagcataaa tcaactggccc aggtgactc agtctgtcca 300
tgtgaaagtg aacttgggta ttccaagtc tatggcttgg tcaatgacaga aggagtgggt 360
tccctgtctt gagagaagac cccgcaagat cctggcctcc ctgcttcaat gtcctgggcc 420
aacaggtgac acaacgttgt gacagctgta ggagcctggc cagctcaatg gacatcctt 480
ggaaatgttc cagaagcccc tgtgggagct gatgtgttgg gggctggagg atgtgactgg 540
gcagacaaag aggccttggc ccttgggaaa agggcaaaag tgcaacttct tcttgagagt 600
tctggacagt ctgatccatc ctatgtctgc ctctctgaca gctgggcaagc cagggagggt 660
ttcccactt acagatctca ggtctcctct ccccgcaccc cgggtagtct catctgggta 720
ggagtggtgt ctggttggcc tatacttggg gaactcaagg aatgtaecc gatgttctcc 780
tgcattgtgc ccactgtgtg tgcctccttc caggatccag gacgttatgg tgattatgac 840
ctccctatgg atgaggatga ggcacatgacc aagcccggga ccttctctgc agccaaatc 900
gtaattggca ttgcatggc aggcacatcag ctggtctgcy gcaatcgtaa ctttctctt 960
atcgctgccc tcaccgccta taagaagttg cgcacactca ccaatctgct cattgccaac 1020
ctggcactct ccgacttctt ggtggcctac atctgtctgc ccttcgagat ggactactac 1080
gtggtacgac agctctcctg ggagcaatggc cactgtctct gtcctccgt caactactct 1140
cgaccctgct cctctactct ctccaccaat gccctgtctg caattgcaat tgacaggtat 1200
ctgcacatct ttaccacctt gaaaccaagc atgaaattac aaacggcctc ctctctgac 1260

WO 02/10387

PCT/US01/23433

```

gccttggctt ggatggtgtc cacttccatt gccatcccat cggcttactt tgcaacagaa 1320
acggctccct ttatgtgcaa gagccaggag aagatcttct gtggccagat ctggcctgtg 1380
gatcagcagc tctactacaa gtccctacttc ctcttccatct btgggtgcga gttcgtgggc 1440
cctgtgggtca ccatgaccct gtgctatgcc aggatctccc gggagctctg gttcaaggca 1500
gtccctgggt tccagacgga gcagattcgc aagcggctgc gctgccgcag gaagcggtc 1560
ctgggtgcca tgtgactctt caaggcctat gtgtgtgtgt gggcaccctt ctaccgtttc 1620
accatcgttc gtgacttctt ccccaactgt ttgtgtgaagg aaaagcacta cctcactgcc 1680
ttctactgtg tctgactgat cgcctagacc aacagctgga tcaacaccgt gtgctctgtg 1740
acggtcaaga acaacaacct gaagtacttc aagaagatga tctgtctgca ctggcgtccc 1800
tcccagcggg ggagcaagtc cagtgtctac ctgtaccctca gaaccaaccg ggtgcccacc 1860
acagaagagg tggactgtat caggctgaag tgaccaactg gtgtcacaca attgaaaaac 1920
ccagtcagat actcagagca tcaaccacca tcaaccaagt tcataggctg catgggaaat 1980
gacatctgtg ttcatgcttc ccccgctgcc tcaagaagcc gaatgtgca aagtctaac 2040
atacaatgag actagacatg aaccaaatca gctgacatct actgatatcc gctcgacacc 2100
tactgtgtcc acaatcccaa caaggagatt agacacaagg agcagcaact gacatggact 2160
gaacatgtac tgtgtgcaag ccaaaccaat gagattaaca gggcagcagc gagctgaatt 2220
atcttactat gtaaccaacc tttgttctac aaatthaaat acagtccaac ttgggtcaca 2280
tcgttttatt tcccatctat tttt 2304

```

- <210> 32
- <211> 2322
- <212> DNA
- <213> Homo sapiens
- <220>
- <221> misc_feature
- <223> Incyte ID No: 7482007CB1

```

<400> 32
cccatttcaa aatggagaa gacagatcac tggcactgac caggaccgtg ggaggtgcca 60
cgtgatgggt agcatcatg ctaggagact gagctctgac ctctctgctg ggtgatctc 120
cacctctggg ctgctgagtc tacttctctg atgccgtgaa gatcctcatg tatgaaaatg 180
aagtcccagg caaccatgat ttgctgctta gtgttcttct tgtccaaga atgttccacc 240
tatagatcca agattaccct aaaagctgga gataaacttc aaagccctga agggaacccc 300
aagactggaa ggtaccaaga gaaatgcgaa ggacctgtta tttcttcttc caactgcagc 360
caaccctgtg ctaaggactt tcatggagaa ataggattta catgtaatca aaaaagttgg 420
caaaaatcag ctgaaacatg tacaagcctt tctgtggaaa aactctttaa ggactcaact 480
gggtcatctc gcccttctgt agcagcacca tctatacctc tgcataatct agactttcga 540
gctccagaga ccattgagag tgtagctcaa ggaatccgta agaactgccc ctttgattat 600
gcctgcatac ctgacatggt gaaatcatca gaaacaacat ctggaaatbt tgcatttata 660
gtgggtttat taaaaaata tttcacagac ttgtctgata atgttactcg agagaaaatg 720
aagagctata gtgagtygc caaccacatc ctgcacacag cagccatttc aaactgggct 780
ttcalttcca acaaaaatgc cagctcggat ttgttgcagt cagtgaattt gtttccaga 840
caactccaca tccacaataa ttctgagaac atgttgaatg aactcttcat tcagacaaaa 900
gggtttccca tcaaccataa tactctcagag aaaagctcca atttctccat ggcgatgaac 960
aataccacag aagatattct agaatggta cagattccca ggcaagactc aaggaaactg 1020
tgcccaaatg catcccaggc cattagcata gctttcccaa ccttgggggc tatcctgaga 1080
gagccaccat tgcaaaatgt gactcttccc agcaggttaa atggtctggt gctatcagt 1140
gttttaccag aaagttgca agaaatcata ctacactctg aaaagatcaa taaaaccctg 1200
aatgcccagag cccagttgtg tggctggcac tccaagaaaa ggagatggga tgaagaagcg 1260
tgccaataga tgttggatat caggaaacgaa gtgaaatgcc gctgtaacta caccagtgtg 1320
ctgatgtctt ttccattctc catgtcctcc aaatcgatga ccgacaaaat tctggactac 1380
atcaactgca ttgggctcag cgtctcaatc ctaagcttgg ttctttgccc gatcattgaa 1440
gccacagtyt ggtcccgggt ggttgtgacg gagatatac acatgcctca cgttgctatc 1500
gtgaatagag cagtgtccc tctgactgcc aatgtgtgt ttatcatagg ctctcacttt 1560
aacatthaaq cccagactca caacatgtgt gttgcagtga catthttcag ccaatttttc 1620
tactctctc tgtttttctg gattctcttc aaagcattgc tcatcattta tggaaatbtg 1680
gtrattttcc gttagatgat gaagtcccga atgatgtca ttggctttgc cattggctat 1740
gggtgccctg tgatctatgc tctcaactca gttgctatca cagggccagt gaaaggctac 1800
atgagacctg aggcctgtgt gcttaactgc gacaatacca aagccctttt agcatttgcc 1860
atcccggcgt tctgctattg ggctgtaaat ctgatgtgg ttttggttgt tctgttcaac 1920
actcagagcc cctctatttg cagttccaag tctcaggatg tggtcataat tatgaggatc 1980
agcaaaaatg ttgccatctc cactccactg ctgggactga cctgggggtt tggaaatagc 2040

```

WO 02/10387

PCT/US01/23433

```

actctcatag aaggaacttc cttgaacttc catabaattt ttgcottgct caatgctttc 2100
caggtttttt tcatctctgct gtttggaaac atlatggatc acaagataag agatgcttgg 2160
agatgtagga tgtcttcaot gaaggggana tggagggcag ctgagaatgc atcaatagc 2220
ccacccaatg gatctaaatt aatgaatcgt caagatgaa atgctgcccc atttctcatg 2280
gatgtcctga gaaccaagag ggagatccag gagaagagg cc 2322

```

```

<210> 33
<211> 2366
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 6769042CB1

```

```

<400> 33
atttagtgga cactatagaa gagccocagt tctggaaag gagatcgcca tgtacttca 60
tgotgcoatt ggaagacatg clttattgtc tcaacgctg ccaagcctct lcatgacatc 120
cacagcaagc cccgtgatgc ccacagatgc ctaccatccc atcataacca acctgacaga 180
agagagaaaa accttccaaa gtcccggagt gatactgagt taactccaaa atgtatccct 240
cagcttaccg agtaagtcct tctcggagca gacagccttg aatctccaaa agacctctt 300
aaaagccgtg ggagagatcc ttctactgcc tggttggatt gctctgtcag aggacagcgc 360
cgtggtactg agtctcatcg acactattga caccgtcatg ggccatgtat cctccaacct 420
gcaacggcagc acgccccagg tcaacgtgga gggctcctct gccatggcag agtttccgt 480
ggccaaaatc ctgcccaga cctggaattc ctcccattac cgtctcccg ccaacgggca 540
gagcttcaic cagatcccc acgagccctt ccacaggtac gcctggagca ccgtcgtggg 600
tctgtgtgac cacagatgc actactacct gaaacaactc tggcccccc accccaagat 660
ccggagagcc atgcatcaac aggactgcct gctgttcgcc accagccacc tgaattccct 720
ggaggtgtcc ccaccacca cctgtctca gaactgtcg ggctctccac tcaataggt 780
ccactcaag cacagattga cagtaagca gccagtgag gccaccaaca gcagcaaccg 840
agtctctgt tactgacct tcttgacct cagctccgga gaaggggtct ggtcgaaca 900
cggctgtgct ctcacagagc gaaacctcac ctactccgtc tgcctgctca ctacctcac 960
caactttgct atctcatgc aggtgggtccc gctggagctt gcaacgggac accaggtggc 1020
gctgtcgtct atcagctatg tgggtgctc cctctccgtc ctctgcctgg tggccaacct 1080
ggtcaccttc gccgtgctgt cctccgtgag caccatccgg aaccagcgtt accaatcca 1140
cgcacaacct tccttgcgct tctgtgtggc ccaggtcctg ctgctcatta gtttccgct 1200
cgagccagc acgaccccc gcccaagtgt ggcctgctc ctacactact tcttctctag 1260
tgccttcgca tggatgctgg tggagggcct gcaactcac agcatggtga tcaaggctct 1320
tgggtccgag gacagcaagc accgttaca ctatgggatg ggaatggggt ttcctctct 1380
gatctgcatc atttcaactg catttgccat ggaagttac ggaacaagca acaatctgt 1440
gctgtcgttg ccgagtgcc ccatctgggc ctctgtgacc cctgcccctg tctctcatct 1500
ggtaaacatt ggcactctca tctgtgtgac cagagtcac tcaagatca cggccgacaa 1560
ctacaagatc catggagacc ccagtcctt caagttgacg gccaaaggcag cggccgtgct 1620
gctgccaatc ctgggtacct cgtgggtctt tggcgtgctt gctgtcaacg gttgtctgt 1680
ggttttccag taactgtttg ccacgtcaca ctccctcag ggaactgtca tcttctctt 1740
tcattgtctc ctgaattcag aggtgagagc cgccttcaag cacaaaatca aggtctgttc 1800
gctcacagag agctccccc gcaactccaa cgcgaagccc tcccaactcg acctcatgaa 1860
tgggacccgg ccaggtatgg cctccaccaa gctcagcctt tgggacaaga gcaaccaatc 1920
tgcccaaccg gtcgacctgt cagccgtgtg agccgggagg ctgccaacca ggcagagctg 1980
cgtcagaac acaccccccc aaacagaatg aaatgcccc cctttgcca tggaccctct 2040
ccttctgctc gtcggacat ggggtgtgtg gccccgagac agctgtctc cctgtgact 2100
ctggctctcg gacgcaactg ctgagccagc cagctgtgat ccagggccag cgtgggacct 2160
cctgcttgc atccaccgtt gggctgagtg acttctcgg gggatccca ggcacagtg 2220
gctgaacttg tgatggtctt cttgagcctc ccttcatcac tcaagatcag accagcgagg 2280
cagggcatcg gggccggtcc cgcagccccg agggatgtca gctctgtgct ggggggttgg 2340
ggcccccccc aagtgtcagg ccccc 2366

```

```

<210> 34
<211> 1458
<212> DNA
<213> Homo sapiens

```

<220>

WO 02/10387

PCT/US01/23433

<221> misc_feature
<223> Incyte ID No: 7476053CB1

<400> 34
atggaggcgc ctageccttc agtggccacc gccggcgttg cccctgcect gggaccgcag 60
accagcagcg gaaccocaaag cccgagaggg atactcgggt cgaccocgag cggcgcgcgc 120
ctcccgggcc gagggccgcc cttctctgtc ttaacggtrc tgggtgtgac gctgctagt 180
ctgctgatgc ctgccacttt cctgtggaac ctgctgttcc oggtoaccaat cccggggctc 240
cgtcccttc acccgcgtgc gcataacttg gtggcctcga cggcgcctc ggaagaaacta 300
gtggcagcgc tggcgtatgc accagccttg gogagtgagc tctgacccgg cgcgctcgg 360
ctgctgggcc ggagcctgtg ccacgtgtgg atctccttgg acgcccctlg ctgcccgcgc 420
ggcctcggga acgtggcgcg catgcacctg ggcgcgcagc gggccatcac acggcaactg 480
cagcacacgc tgcgcaccgc cagccgcgcc tctgtgtcca tggatcgctc cgcccgggtg 540
ccgtcggcgc tcatgccctt cgcgccgctg ctctttggcc ggggcgaggt gtgcgacgtc 600
cggctccagc gctgccaggt gagccgggaa cctcctatg ccgccccttc caccgcggcc 660
cgcctccacc tgcgccttgg cgtgggtgcc tttgtctacc ggaagatcta cgaaggcggcc 720
aagtttcggt tggcgcgcg cgggagagct gtcctgcccgt tgcggccac catgcagggt 780
aaggagcac ctgatgagcg tgaagtgttg ttcacggcac ctgtcaaaag aacggtgtcc 840
tccagggtga gccgggactc ctggcggagc cagnaagaga ggcgagcagc catgatggtg 900
ggaattctga ttggcgtgtt tgtctgtgc tggatccctt tcttctgac ggaactcctc 960
agcccaactc tggcctcagc cctgcccccc atctggaaaa gcatatttct gtggcttggc 1020
tactccaatt ctttcttcaa ccccctgatt tacacagctt taaacaagaa ctacaacaat 1080
gcctccaaga gcctctttac taagcagaga tgaacacagg ggttagagag acatgggtag 1140
atthtaagga ggaaggaact tggacttttt cgtcagtgat ctgagattct tccctccaca 1200
gctgagtgct aatgctgtat tgagagttat accattgggc ctggactgta gaagcagcag 1260
agccaaagtt ctaagaagaag acagcaaaag tctggcagat tctgtaacta tgccttcttc 1320
ccatgtcatt ggcagcatt gccaatgggt catggcttgg ctcccactg agcaggaact 1380
tggctcagca atcctttcca ggacagcacc ctaggcagct actgttgatt atttaaaatt 1440
gatgcaagac tlgaaaaa 1458

<210> 35
<211> 975
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7480410CB1

<400> 35
atgggcatgy agggctctct ccagaactcc actaactcgg tccctcacagg cctcatcacc 60
catctcgcct tcccggggt tctctttgca atagtctctt ccaatcttgt ggtggctata 120
acagccaact tggctcagat tctgtctatc cacatggact cccgctcca cactcccatg 180
tactctttgc tcagccagct ctccatcatg gataccatct acatctgtat cactgtcccc 240
aagatgtctc aggaccctct gtccaagcac aagaccattt ccttctctgg ctgtgcagtt 300
cagatcttcc tctaccctgac cctgatgga ggggaattct tectgtctgg tctcatggcc 360
tatgacgcgt atgtggctgt gtgaacctc ctacggtacc ctctctcat gaaccgcagg 420
gtttgcttat tcatgggtgt cggctcctgg gttgggtggt ccttggatgg gttcatgctg 480
actcctgcta ctatgagttt cccctctgt agatcccagc agatcaatca ctttttctgt 540
gagatcccag ccgtgctgaa gttgtcttgc acagacacgt cactctatga gaccctgatg 600
tatgctgctc cgtgtgctat gctgcttacc cctctatctg tcatctctgt ctccatcacg 660
cacatcctcc tgaactgtcca caggatgaac tctgctgagg gccggcgcaa agcctttgct 720
acgtgtctcc cccacattat ggtgggtgagc gttttctacg gggcagcctt ctacaccaac 780
gtgctgccc actcctacca cactccagag aaagataaag tggatpctgc ctctcaacc 840
atctccacc ccatgctcaa cccactcctc tacagcttga ggaataaaga tgtggctgca 900
gctctgagga aagtactagg gagatgtggt tctccacaga gcatcagggt ggcgactgtg 960
atcaggaagg gctag 975

<210> 36
<211> 948
<212> DNA
<213> Homo sapiens

WO 02/10387

PCT/US01/23433

<220>
<221> misc_feature
<223> Incyte ID No: 55036418CB1

<400> 36
atggagacgt ggtgaacca gtactacaca gatggcttct tctcttagg catcttctcc 60
caacgtactg ctgaacctgt cctcttctcc g'gggtatgg cggtrctcac agtggccctc 120
tgtgggaatg tctctctcat ctctctcacc tacatggacc ctacacctca caccoccatg 180
tacttcttcc tcagccagct ctccctcatg gacctcatgt tggctgttac caatgtgcca 240
aagatggcag ccaactctct gtctggcagg aagtccatct cctttgtggg ctgtggcaca 300
caaatggcc tcctttgtctg tcttgggga tctgaggggc tcttctggg actcatggct 360
tatgaccgct atgtggccat tagcaccaca ctctcactac ccatcctcat gaatcagagg 420
gtctgtctcc agattactgg gagctcctgg gcctttggga taatcgatgg cttgatccag 480
atggtggtag taatgaattt cccctactgt ggttgaggga agtgaacca ttttctctgt 540
ggatgctat ccttgtttaa gctggcctgt gtagacacat ccctgtttga gaagtgata 600
tttggctgct gtgtcttcat gcttctcttc ccattctcca tcatctgggc ctccatgctt 660
ccatttctag ggaactgtct gcaaatgac tctgtctcagg cctggaaaaa ggccttggcc 720
acctgctctc ccacactgac agctgtcacc ctctctctatg gggcagccat gttcatctac 780
ctgagacctc ggcactacc ggccccacc catgacaagg tggcctctat ctctcaccag 840
gtccttactc ccactgctca cccctcatt tacagcttga ggaacaggga ggtgatgggg 900
gcactgagga aggggctgga ccgctgcagg atcggcagcc agcactga 948

<210> 37
<211> 1086
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481701CB1

<400> 37
ggctctatcc agacgctggc ttcttgttag tgtattcctt tatccaatag tagatgctcc 60
ctagaaggct tgagtgcact ggaattttaa actctcactt tcaacttga gatggagagc 120
cccactcaaa ccaccattca ggagtttacc ttctccgctt tcccttacc ctgggttaag 180
tctgtgtctc gctttgttcc actgctcttc atctatgctt tcattgttgt tggaaacctg 240
gtcatcaca cagtgggtcca gttgaatact caccctccaca ctcccatgta tacttttacc 300
agtgtccttt cttttctgga gatttggatc accacagcca caatcccaaa gatgtgtct 360
agctgcttca gtgagaggag catttccctc aatgggtgtc tctcgagat gtatttcttc 420
cattccaccg gcactctgta ggtgtgtctc ttgacagtta tggccttga ccactacctg 480
gcaatagca gccctcttca ttatccctct atcatgacc ccaagctatg taccacaactg 540
actttaagtt gctgtgttgg tggctttacc acacccttcc ctgagatgce ctggatctct 600
acactgccat ttgtgtgttc gaatcaccctt gaacatctct tctgtgactt cctccagctg 660
ctgctgtctg cctgcacaga cacacagacc atcgtcatga ttcaggtagt ggatgtcatt 720
catgcaatgg agattattac agctgtgatg ctcatctcca tctcctcaga tggatattgt 780
gctgtaatca tacgtattca ttcagctgga ggcggccgca cagcatttcc cactgtgttc 840
tctcacttca ttgtcttttc getctctctt ggcagtgta ctctcatgta cctacgcttc 900
tctgcccact actctttgtt ctgggatata gccattgctc tggcctttgc agttttgtct 960
cccttcttca accccattat ctatagcctg aggaataaag aaataaaga agctataaaa 1020
aagcacatag gtcagctaa gatatttttt tccgtaagac cagggaacctc aagtaagata 1080
ttttag 1086

<210> 38
<211> 1529
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481774CB1

<400> 38
aagggagacc acagtgagag ggagccctga gcagaagtaa ggctgtcaca aggctggaag 60

WO 02/10387

PCT/US01/23433

```

cagagaacat ccccatggaa ctgaagacag catgctgcat ccoctgggagg agggagctct 120
caaggaagtt ccaaggattg atatttctgt tcagctgcag tagagatgga tggaaacctg 180
gtagaaccca acccatctca tcctactggg attctctgac cgaccccac tcggagaggat 240
cctctttgtg gtatctctga tgggtacct cctgacctc gtaggcaaca ccaccatcat 300
cctgggtgce eggctggaac cccacctcca caccoccatg tacctctctc tggcccaect 360
ttcctctctg gacctcagtt tcaccaccag ctccatcccc cagctgctct acaaccttaa 420
tggatgtgac aagaccatca gctacatggg ctgtgccalc cagctcttcc tgttctggg 480
tctgggtggc gtggagtgc tcctctggc tgcctggcc tatgaocgt gtgtggctat 540
ctgcaagccc ctgcactaca tgggtgcat gaaccccagg ctctgccgg gcttgggtc 600
agtgaactgg ggctgtgggg tggccaactc cttggccatg tctctctgta ccttgggct 660
accccgtgt ggccaccacg aggtggacca cttcctgtgt gagatgccc cctgatccg 720
gatggcctgc atcagcactg tggccatcga cggcaccgtc ttgtctctg cgttgggtgt 780
tgtgctgccc ccttgggtt ttatcctgct ctcttacagc tacattgtga gggctgtgt 840
acaaattcgg tcagcatcag gaaggcagaa ggcttcggc acctgggct cccatctca 900
tgtggctccc ctttctatg gaaacatcat ctacatgtac atgcagccag gagccagtcc 960
ttcccaggac cagggcaagt tcctcacgct cttctacaac atgtctcccc cctcctcaa 1020
tcctctcabc tacacctca gaaacagaga ggtgagggg gcactgggaa ggttctctc 1080
ggggaagaga gagctaggaa aggagtaag gcatctcac ctgacttcac ctccatccag 1140
ggccactggc agcatctgga acgctgat tcacgtgat attagccac gactcccaa 1200
ttgcttttt ctggactttt gtgagctgt ttcagttctg acattatgt tttttgtgt 1260
tgctctaaa atgagacgg ggtctcactc tgcacttag ggtggatgc agtgggtcca 1320
ccatagctcc ttgactatt gggcttaagc gatcctccc cacctcagcc tccaagtaa 1380
ctgggactac aggtgtgcat cactgycagt gggaaatgtg gctttctgt ettcattgga 1440
gacggggctc tgcctgttg accaggctgg tcccaactc ctggcctcat gtgatctcc 1500
tgccatggcc tcctaaagtt ctgggatta 1529

```

【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
7 February 2002 (07.02.2002)

PCT

(10) International Publication Number
WO 02/010387 A3(51) International Patent Classification: C12N 15/12,
15/63, 1/21, 5/10, A01K 67/027, C07K 14/705, 16/28,
C12Q 1/68, G01N 33/50, 33/577, A61K 38/17, 39/395

(21) International Application Number: PCT/US01/23433

(22) International Filing Date: 25 July 2001 (25.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/221,478 27 July 2000 (27.07.2000) US
60/223,268 3 August 2000 (03.08.2000) US
60/227,054 21 August 2000 (21.08.2000) US
60/231,121 8 September 2000 (08.09.2000) US
60/232,243 13 September 2000 (13.09.2000) US
60/232,691 15 September 2000 (15.09.2000) US
60/235,146 22 September 2000 (22.09.2000) 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): THORNTON,
Michael [US/US]; 9 Medway Road, Woodside, CA 94062
(US). ARVIZU, Chandra [US/US]; 1706 Morocco Drive,
San Jose, CA 95125 (US). LAL, Preeti [IN/US]; P.O.
Box 5142, Santa Clara, CA 95056 (US). BURFORD,
Neil [GB/US]; 105 Wildwood Circle, Durham, CT 06422
(US). YUE, Henry [US/US]; 826 Lois Avenue, Sunny-
vale, CA 94087 (US). GANDHI, Ameena, R. [US/US];
837 Roble Avenue, #1, Redwood City, CA 94025 (US).
ELLIOT, Vicki, S. [US/US]; 3770 Pollen Place Way,
San Jose, CA 95121 (US). RAMKUMAR, Jayalaxmi
[IN/US]; 34359 Maybied Circle, Fremont, CA 94555 (US).
BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road,
San Leandro, CA 94577 (US). KALLICK, Deborah,
A. [US/US]; 58 Linda Vista, Atherton, CA 94027 (US).
WALLA, Narinder, K. [US/US]; 890 Davis Street #205,
San Leandro, CA 94577 (US). HAFALIA, April, J.,A. [US/US]; 2227 Calle de Primavera, Santa Clara, CA
95054 (US). YAO, Monique, G. [US/US]; 1189 Woodgate
Drive, Carmel, IN 46033 (US). LU, Yan [CN/US]; 3885
Corrina Way, Palo Alto, CA 94303 (US). TRIBOULEY,
Catherine, M. [FR/US]; 1121 Tennessee Street, #5,
San Francisco, CA 94107 (US). POLICKY, Jennifer,
L. [US/US]; 1511 Jarvis Court, San Jose, CA 95118
(US). KEARNEY, Liam [IE/US]; 50 Woodside Avenue,
San Francisco, CA 94127 (US). GRAUL, Richard, C.
[—/US]; 682-29th Avenue, San Francisco, CA 94121
(US). WARREN, Bridget, A. [US/US]; 2250 Homestead
Court, #2, Los Altos, CA 94024 (US). LEE, Ernestine, A.
[US/US]; 3910 Fairroaks Avenue, Menlo Park, CA 94025
(US). DING, Li [US/US]; 3353 Alma Street, #146, Palo
Alto, CA 94306 (US).(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics,
Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).(81) Designated States (national): AT, AG, AL, AM, 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, GI, GM,
GR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,
LR, LS, LU, LV, MA, MD, MG, MK, MN, MW, MX,
MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL,
TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG).

Published:

— with international search report

(88) Date of publication of the international search report:
3 January 2005For 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 02/010387 A3

(54) Title: G-PROTEIN COUPLED RECEPTORS

(57) Abstract: The invention provides human G-protein coupled receptors (GCR1C) and polynucleotides which identify and encode GCR1C. The invention also provides expression vectors, host cells, antibodies, agonist, and antagonist. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of GCR1C.

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT		International Application No PCT/US 01/23433
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/12 C12N15/63 C12N1/21 C12N5/10 A01K67/027 C07K14/705 C07K16/28 C12Q1/68 G01N33/50 G01N33/577 A61K38/17 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 A01K C07K C12Q G01N A61K		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
EMBL, SEQUENCE SEARCH, EPO-Internal, WPI Data, PAJ, BIOSIS, MEDLINE, CHEM ABS Dat		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] accession: AC026960, 27 March 2000 (2000-03-27) BIRREN B ET AL: "Homo sapiens chromosome 3 clone RP11-170K4 map 3, WORKING DRAFT SEQUENCE, 15 unordered pieces." XP002195514	1-8, 11, 12, 45, 64
P,X	WO 01 48188 A (MATSUMOTO SHUN ICHIRO ; MORIKAWA NORIYUKI (JP); SUGIYAMA TOMOYASU () 5 July 2001 (2001-07-05) SEQ ID NO: 20; SEQ ID NO: 25 abstract; claims 1-17; example 1 -/-	1-19, 22, 25-45, 64
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "Z" document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
9 April 2002		23. 07. 2002
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2200 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl. Fax: (+31-70) 340-3016		Authorized officer Devijver, K

Form PCT/ISA/210 (second sheet) (July 1999)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/23433

C/(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category ¹	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 01 36473 A (PARODI LUIS A ;LIND PETER (SE); SEJLITZ TORSTEN (SE); SCHELLIN KAT) 25 May 2001 (2001-05-25) SEQ ID NO: 13; SEQ ID NO: 14 claims 1-97; examples 1-15 ---	1-19,22, 25-45,64
E	WO 01 85764 A (BAYER AG ;RAMAKRISHNAN SHYAM (US)) 15 November 2001 (2001-11-15) SEQ ID NO: 1 claims 1-78; examples 1-14 ---	1-19,22, 25-45,64
E	WO 01 66750 A (UPJOHN CO ;VOGELI GABRIEL (US); WOOD LINDA S (US)) 13 September 2001 (2001-09-13) SEQ ID NO: 47; SEQ ID NO: 181 abstract; claims 1-95 ---	12
E	WO 01 62797 A (PARODI LUIS A ;LIND PETER (SE); UPJOHN CO (US); VOGELI GABRIEL (US)) 30 August 2001 (2001-08-30) SEQ ID NO: 32 abstract; claims 1-141 ---	12
A	STAM N J ET AL: "MOLECULAR CLONING AND CHARACTERIZATION OF A NOVEL ORPHAN RECEPTOR EXPRESSED IN HUMAN PANCREAS THAT SHOWS HIGH STRUCTURAL HOMOLOGY TO THE P2U PURINOCEPTOR" FEBS LETTERS, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 384, no. 3, 22 April 1996 (1996-04-22), pages 260-264, XP002030124 ISSN: 0014-5793 ---	
A	PARR C E ET AL: "CLONING AND EXPRESSION OF A HUMAN P2U NUCLEOTIDE RECEPTOR, A TARGETFOR CYSTIC FIBROSIS PHARMACOTHERAPY" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE, WASHINGTON, US, vol. 91, 1 April 1994 (1994-04-01), pages 3275-3279, XP000611412 ISSN: 0027-8424 ---	
A	WO 99 55732 A (AHMAD SULTAN ;CAO JACK (CA); DONNELL DAJAN O (CA); WALKER PHILIPPE) 4 November 1999 (1999-11-04) -----	

1

Form PCT/ISA210 (continuation of second sheet) (July 1993)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 01/23433

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. 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. 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. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 5.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-82 (all partially)

Remark on Protest

- The additional search fees were accompanied by the applicant's protest.
 No protest accompanied the payment of additional search fees.

International Application No. PCT/US 01/23433

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.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 claims 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 I.2

Claims Nos.: 20,21,23,24

Claim 20 refers to an agonist compound of a polypeptide of claim 1 identified by a method of claim 19 without giving a true technical characterization. Moreover, no such specific compounds are defined in the application. In consequence, the scope of said claim is ambiguous and vague, and its subject-matter is not sufficiently disclosed and supported (Art. 5 and 6 PCT).

No meaningful search can be carried out for such purely speculative claims whose wording is, in fact, a mere recitation of the result to be achieved.

The above comment also applies for claims 21, 23 and 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 II procedure.

International Application No. PCT/US 01/23433

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

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

Invention 1: claims: in part: 1-82; all as far as applicable

An isolated polypeptide and polynucleotide relating to SEQ ID NOS 1 and 20, and fragments and variants thereof. Expression vector, host cell and transgenic organism comprising such a polynucleotide. Method for producing such a polypeptide. Antibody which specifically binds to such a polypeptide. Method for detecting such a polynucleotide or such a polypeptide. Composition comprising such a polypeptide or such an antibody. Methods for screening agonists or antagonists of such a polypeptide. Composition comprising such agonists or antagonists. Method of treatment by using such compositions. Method for screening for a compound that specifically binds to such a polypeptide or that modulates the activity of such a polypeptide. Method for screening a compound for effectiveness in altering expression of such a polynucleotide. Method for assessing toxicity of a test compound by using such a polynucleotide. A diagnostic test and method of diagnosis using such an antibody. Method for preparing such a polyclonal or monoclonal antibody. Method of purifying such a polypeptide using such an antibody.

Inventions 2-19: claims: in part: 1-82; all as far as applicable

As invention 1, but limited to subject-matter relating to SEQ ID NOS 2-19 and 21-38, wherein:
invention 2 is limited to SEQ ID NOS 2 and 21,
invention 3 is limited to SEQ ID NOS 3 and 22,
....
invention 19 is limited to SEQ ID NOS 19 and 38.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 01/23433

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 0148188	A	05-07-2001	AU 2230301 A	09-07-2001
			AU 2230401 A	09-07-2001
			WO 0148188 A1	05-07-2001
			WO 0148189 A1	05-07-2001
			AU 4467301 A	08-10-2001
			WO 0173023 A1	04-10-2001
WO 0136473	A	25-05-2001	AU 1617801 A	30-05-2001
			WO 0136473 A2	25-05-2001
WO 0185764	A	15-11-2001	AU 7240001 A	20-11-2001
			WO 0185764 A2	15-11-2001
WO 0166750	A	13-09-2001	AU 4009701 A	17-09-2001
			WO 0166750 A2	13-09-2001
WO 0162797	A	30-08-2001	AU 4165801 A	03-09-2001
			AU 5787501 A	09-07-2001
			WO 0148015 A2	05-07-2001
			WO 0162797 A2	30-08-2001
			US 2002062013 A1	23-05-2002
			AU 4165501 A	03-09-2001
WO 9955732	A	04-11-1999	AU 4298099 A	16-11-1999
			CA 2325869 A1	04-11-1999
			EP 1071714 A1	31-01-2001
			JP 2002512779 T	08-05-2002
			WO 9955732 A1	04-11-1999

フロントページの続き

(51) Int.Cl. ⁷	F I		テーマコード(参考)
A 6 1 P 1/16	A 6 1 P 1/18		4 C 0 8 4
A 6 1 P 1/18	A 6 1 P 3/00		4 C 0 8 5
A 6 1 P 3/00	A 6 1 P 3/10		4 H 0 4 5
A 6 1 P 3/10	A 6 1 P 5/14		
A 6 1 P 5/14	A 6 1 P 7/00		
A 6 1 P 7/00	A 6 1 P 7/04		
A 6 1 P 7/04	A 6 1 P 7/06		
A 6 1 P 7/06	A 6 1 P 9/04		
A 6 1 P 9/04	A 6 1 P 9/10		
A 6 1 P 9/10	A 6 1 P 9/10	1 0 1	
A 6 1 P 9/12	A 6 1 P 9/12		
A 6 1 P 11/06	A 6 1 P 11/06		
A 6 1 P 13/02	A 6 1 P 13/02		
A 6 1 P 17/06	A 6 1 P 17/06		
A 6 1 P 19/00	A 6 1 P 19/00		
A 6 1 P 19/06	A 6 1 P 19/06		
A 6 1 P 19/10	A 6 1 P 19/10		
A 6 1 P 21/00	A 6 1 P 21/00		
A 6 1 P 21/04	A 6 1 P 21/04		
A 6 1 P 25/00	A 6 1 P 25/00		
A 6 1 P 25/02	A 6 1 P 25/02	1 0 3	
A 6 1 P 25/08	A 6 1 P 25/08		
A 6 1 P 25/14	A 6 1 P 25/14		
A 6 1 P 25/16	A 6 1 P 25/16		
A 6 1 P 25/18	A 6 1 P 25/18		
A 6 1 P 25/20	A 6 1 P 25/20		
A 6 1 P 25/22	A 6 1 P 25/22		
A 6 1 P 25/28	A 6 1 P 25/28		
A 6 1 P 29/00	A 6 1 P 29/00		
A 6 1 P 31/04	A 6 1 P 29/00	1 0 1	
A 6 1 P 31/10	A 6 1 P 31/04		
A 6 1 P 31/12	A 6 1 P 31/10		
A 6 1 P 35/00	A 6 1 P 31/12		
A 6 1 P 35/02	A 6 1 P 35/00		
A 6 1 P 37/02	A 6 1 P 35/02		
A 6 1 P 37/08	A 6 1 P 37/02		
A 6 1 P 43/00	A 6 1 P 37/08		
C 0 7 K 1/22	A 6 1 P 43/00	1 0 5	
C 0 7 K 14/705	A 6 1 P 43/00	1 1 1	
C 0 7 K 16/28	C 0 7 K 1/22		
C 1 2 N 1/15	C 0 7 K 14/705		
C 1 2 N 1/19	C 0 7 K 16/28		
C 1 2 N 1/21	C 1 2 N 1/15		
C 1 2 N 5/10	C 1 2 N 1/19		
C 1 2 P 21/02	C 1 2 N 1/21		
C 1 2 P 21/08	C 1 2 P 21/02		C
C 1 2 Q 1/68	C 1 2 P 21/08		

G 0 1 N 33/15	C 1 2 Q 1/68	A
G 0 1 N 33/50	G 0 1 N 33/15	Z
G 0 1 N 33/53	G 0 1 N 33/50	Z
G 0 1 N 33/566	G 0 1 N 33/53	D
	G 0 1 N 33/53	M
	G 0 1 N 33/566	
	C 1 2 N 5/00	A
	A 6 1 K 37/02	

- (31)優先権主張番号 60/231,121
 (32)優先日 平成12年9月8日(2000.9.8)
 (33)優先権主張国 米国(US)
 (31)優先権主張番号 60/232,243
 (32)優先日 平成12年9月13日(2000.9.13)
 (33)優先権主張国 米国(US)
 (31)優先権主張番号 60/232,691
 (32)優先日 平成12年9月15日(2000.9.15)
 (33)優先権主張国 米国(US)
 (31)優先権主張番号 60/235,146
 (32)優先日 平成12年9月22日(2000.9.22)
 (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,GQ,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,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,S D,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZW

- (72)発明者 アービズ、チャンドラ
 アメリカ合衆国カリフォルニア州9 5 1 2 5・サンノゼ・モロッコドライブ 1 7 0 6
 (72)発明者 ラル、ブリーティ
 アメリカ合衆国カリフォルニア州9 5 0 5 6・サンタクララ・ピーオーボックス 5 1 4 2
 (72)発明者 パーフォード、ニール
 アメリカ合衆国コネチカット州0 6 4 2 2・ダラム・ワイルドウッドサークル 1 0 5
 (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 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・レッドウッドシティー・# 1・ローブルアベニュー
 8 3 7
 (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 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 ボーゲン、マライア・アール
 アメリカ合衆国カリフォルニア州9 4 5 7 7・サニーベイル・サンティアゴロード 1 4 2 4 4
 (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 7・アサートン・リンダピスタ 5 8
 (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 ハファリア、エープリル・ジェイ・エイ
アメリカ合衆国カリフォルニア州 9 5 0 5 4 ・ サンタクララ ・ コーレデブリマベータ 2 2 2 7
- (72)発明者 ヤオ、モニック・ジー
アメリカ合衆国インディアナ州 4 6 0 3 3 ・ カメル・ウッドゲートドライブ 1 1 8 9
- (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 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 ポリッキー、ジェニファー・エル
アメリカ合衆国カリフォルニア州 9 5 1 1 8 ・ サンノゼ ・ ジャービスコート 1 5 1 1
- (72)発明者 キーニー、ライアム
アメリカ合衆国カリフォルニア州 9 4 1 2 7 ・ サンフランシスコ ・ ウッドサイドアベニュー 5 0
- (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 ・ トウエンティナインスアベニュー
6 8 2
- (72)発明者 ワレン、ブリジット、エイ
アメリカ合衆国カリフォルニア州 9 4 0 2 4 ・ ロスアルトス ・ # 2 ・ ホームステッドコート 2 2
5 0
- (72)発明者 リー、アーンステーション・エイ
アメリカ合衆国カリフォルニア州 9 4 0 2 5 ・ メンロパーク ・ フェアオークスアベニュー 3 9 1
0
- (72)発明者 ディング、リー
アメリカ合衆国カリフォルニア州 9 4 3 0 6 ・ パロアルト ・ # 1 4 6 ・ アルマストリート 3 3 5
3

F ターム(参考) 2G045 AA25 AA40 BB03 BB20 CB01 CB21 DA12 DA13 DA14 DA36
DA37 DA77 FB02 FB03
4B024 AA01 AA11 BA43 BA61 BA63 CA04 DA01 DA02 DA05 DA06
DA11 EA04
4B063 QA01 QA18 QQ02 QQ08 QQ42 QQ53 QQ61 QR08 QR32 QR42
QR50 QR55 QR72 QR77 QS03 QS24 QS25 QS34
4B064 AG20 AG27 CA02 CA05 CA10 CA11 CA19 CA20 CC24 DA01
4B065 AA01X AA57X AA87X AA90Y AB01 AB02 BA02 BA08 CA24 CA25
CA44 CA46
4C084 AA02 AA06 AA17 BA01 BA22 CA53 DC50 NA14 ZA012 ZA022
ZA052 ZA062 ZA082 ZA152 ZA162 ZA182 ZA202 ZA362 ZA392 ZA402
ZA422 ZA452 ZA532 ZA592 ZA662 ZA682 ZA732 ZA752 ZA812 ZA892
ZA942 ZA962 ZA972 ZB052 ZB072 ZB112 ZB132 ZB152 ZB212 ZB262
ZB272 ZB332 ZB352 ZC022 ZC062 ZC212 ZC352
4C085 AA13 BB11 CC04 DD33 EE01
4H045 AA10 AA11 AA20 AA30 BA09 CA40 DA50 DA75 DA76 EA20
EA50 FA74

专利名称(译)	G蛋白偶联受体		
公开(公告)号	JP2004516817A	公开(公告)日	2004-06-10
申请号	JP2002516303	申请日	2001-07-25
[标]申请(专利权)人(译)	洞察Genomics公司		
申请(专利权)人(译)	洞察基因组公司		
[标]发明人	ソーントンマイケル アービズチャンドラ ラルプリーティ バーフォードニール ユエヘンリー ガンディーアミーナアール エリオットビッキーエス ランクマールジャヤラクシミ ボーグンマライアアール カリックデボラーエイ チョーラナリンダーケイ ハファリアエープリルジェイエイ ヤオモニークジー リュヤン トリボレーキャサリーンエム ポリッキージェニファーエル キーニーライアム グラールリチャード ワレンブリジットエイ リーアーンステイーンエイ デイングリー		
发明人	ソーントン、マイケル アービズ、チャンドラ ラル、プリーティ バーフォード、ニール ユエ、ヘンリー ガンディー、アミーナ・アール エリオット、ビッキー・エス ランクマール、ジャヤラクシミ ボーグン、マライア・アール カリック、デボラー・エイ チョーラ、ナリンダー・ケイ ハファリア、エープリル・ジェイ・エイ ヤオ、モニーク・ジー リュ、ヤン トリボレー、キャサリーン・エム ポリッキー、ジェニファー・エル キーニー、ライアム グラール、リチャード ワレン、ブリジット、エイ リー、アーンステイーン・エイ デイング、リー		

IPC分类号	G01N33/50 A61K38/00 A61K38/17 A61K39/395 A61K45/00 A61K48/00 A61P1/04 A61P1/16 A61P1/18 A61P3/00 A61P3/10 A61P5/14 A61P7/00 A61P7/04 A61P7/06 A61P9/04 A61P9/10 A61P9/12 A61P11/06 A61P13/02 A61P17/06 A61P19/00 A61P19/06 A61P19/10 A61P21/00 A61P21/04 A61P25/00 A61P25/02 A61P25/08 A61P25/14 A61P25/16 A61P25/18 A61P25/20 A61P25/22 A61P25/28 A61P29/00 A61P31/04 A61P31/10 A61P31/12 A61P35/00 A61P35/02 A61P37/02 A61P37/08 A61P43/00 C07K1/22 C07K14/705 C07K16/28 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N15/09 C12N15/12 C12P21/02 C12P21/08 C12Q1/68 G01N33/15 G01N33/53 G01N33/566 G01N33/74
CPC分类号	A01K2217/05 A61K38/00 A61K48/00 A61P1/04 A61P1/16 A61P1/18 A61P3/00 A61P3/10 A61P5/14 A61P7/00 A61P7/04 A61P7/06 A61P9/04 A61P9/10 A61P9/12 A61P11/06 A61P13/02 A61P17/06 A61P19/00 A61P19/06 A61P19/10 A61P21/00 A61P21/04 A61P25/00 A61P25/02 A61P25/08 A61P25/14 A61P25/16 A61P25/18 A61P25/20 A61P25/22 A61P25/28 A61P29/00 A61P31/04 A61P31/10 A61P31/12 A61P35/00 A61P35/02 A61P37/02 A61P37/08 A61P43/00 C07K14/705 G01N33/74 G01N2333/726 G01N2500/00
FI分类号	C12N15/00.ZNA.A A61K39/395.D A61K45/00 A61P1/04 A61P1/16 A61P1/18 A61P3/00 A61P3/10 A61P5/14 A61P7/00 A61P7/04 A61P7/06 A61P9/04 A61P9/10 A61P9/10.101 A61P9/12 A61P11/06 A61P13/02 A61P17/06 A61P19/00 A61P19/06 A61P19/10 A61P21/00 A61P21/04 A61P25/00 A61P25/02.103 A61P25/08 A61P25/14 A61P25/16 A61P25/18 A61P25/20 A61P25/22 A61P25/28 A61P29/00 A61P29/00.101 A61P31/04 A61P31/10 A61P31/12 A61P35/00 A61P35/02 A61P37/02 A61P37/08 A61P43/00.105 A61P43/00.111 C07K1/22 C07K14/705 C07K16/28 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/AA25 2G045/AA40 2G045/BB03 2G045/BB20 2G045/CB01 2G045/CB21 2G045/DA12 2G045/DA13 2G045/DA14 2G045/DA36 2G045/DA37 2G045/DA77 2G045/FB02 2G045/FB03 4B024/AA01 4B024/AA11 4B024/BA43 4B024/BA61 4B024/BA63 4B024/CA04 4B024/DA01 4B024/DA02 4B024/DA05 4B024/DA06 4B024/DA11 4B024/EA04 4B063/QA01 4B063/QA18 4B063/QQ02 4B063/QQ08 4B063/QQ42 4B063/QQ53 4B063/QQ61 4B063/QR08 4B063/QR32 4B063/QR42 4B063/QR50 4B063/QR55 4B063/QR72 4B063/QR77 4B063/QS03 4B063/QS24 4B063/QS25 4B063/QS34 4B064/AG20 4B064/AG27 4B064/CA02 4B064/CA05 4B064/CA10 4B064/CA11 4B064/CA19 4B064/CA20 4B064/CC24 4B064/DA01 4B065/AA01X 4B065/AA57X 4B065/AA87X 4B065/AA90Y 4B065/AB01 4B065/AB02 4B065/BA02 4B065/BA08 4B065/CA24 4B065/CA25 4B065/CA44 4B065/CA46 4C084/AA02 4C084/AA06 4C084/AA17 4C084/BA01 4C084/BA22 4C084/CA53 4C084/DC50 4C084/NA14 4C084/ZA012 4C084/ZA022 4C084/ZA052 4C084/ZA062 4C084/ZA082 4C084/ZA152 4C084/ZA162 4C084/ZA182 4C084/ZA202 4C084/ZA362 4C084/ZA392 4C084/ZA402 4C084/ZA422 4C084/ZA452 4C084/ZA532 4C084/ZA592 4C084/ZA662 4C084/ZA682 4C084/ZA732 4C084/ZA752 4C084/ZA812 4C084/ZA892 4C084/ZA942 4C084/ZA962 4C084/ZA972 4C084/ZB052 4C084/ZB072 4C084/ZB112 4C084/ZB132 4C084/ZB152 4C084/ZB212 4C084/ZB262 4C084/ZB272 4C084/ZB332 4C084/ZB352 4C084/ZC022 4C084/ZC062 4C084/ZC212 4C084/ZC352 4C085/AA13 4C085/BB11 4C085/CC04 4C085/DD33 4C085/EE01 4H045/AA10 4H045/AA11 4H045/AA20 4H045/AA30 4H045/BA09 4H045/CA40 4H045/DA50 4H045/DA75 4H045/DA76 4H045/EA20 4H045/EA50 4H045/FA74
优先权	60/221478 2000-07-27 US 60/223268 2000-08-03 US 60/227054 2000-08-21 US 60/231121 2000-09-08 US 60/232243 2000-09-13 US 60/232691 2000-09-15 US 60/235146 2000-09-22 US
外部链接	Espacenet

摘要(译)

本发明提供了人G蛋白偶联受体 (GCREC) 和鉴定和编码GCREC的多核苷酸。本发明还提供了表达载体, 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的异常表达相关的病症的方法。

