

(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2004-533222

(P2004-533222A)

(43) 公表日 平成16年11月4日(2004.11.4)

(51) Int. Cl. ⁷	F I	テーマコード (参考)
C 1 2 N 15/09	C 1 2 N 15/00	Z N A A
A 6 1 K 38/00	A 6 1 K 45/00	2 G O 4 5
A 6 1 K 45/00	A 6 1 K 49/00	4 B O 2 4
A 6 1 K 49/00	A 6 1 P 3/10	4 B O 2 9
A 6 1 P 3/10	A 6 1 P 7/06	4 B O 6 3
		4 B O 6 4
審査請求 未請求 予備審査請求 未請求 (全 245 頁) 最終頁に続く		

(21) 出願番号	特願2002-571850 (P2002-571850)	(71) 出願人	301005050
(86) (22) 出願日	平成14年3月12日 (2002.3.12)		インサイト・ゲノミックス・インコーポレイテッド
(85) 翻訳文提出日	平成15年9月9日 (2003.9.9)		アメリカ合衆国カリフォルニア州94304・パロアルト・ポータードライブ 3160
(86) 国際出願番号	PCT/US2002/009052	(74) 代理人	100089266
(87) 国際公開番号	W02002/072794		弁理士 大島 陽一
(87) 国際公開日	平成14年9月19日 (2002.9.19)	(72) 発明者	ユエ、ヘンリー
(31) 優先権主張番号	60/275, 249		アメリカ合衆国カリフォルニア州94087・サニーベイル・ルイスアベニュー 826
(32) 優先日	平成13年3月12日 (2001.3.12)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/316, 810		
(32) 優先日	平成13年8月31日 (2001.8.31)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/323, 977		
(32) 優先日	平成13年9月21日 (2001.9.21)		
(33) 優先権主張国	米国 (US)		
最終頁に続く			

(54) 【発明の名称】 免疫グロブリンスーパーファミリータンパク質

(57) 【要約】

本発明はヒトの免疫グロブリンスーパーファミリータンパク質 (IGSFP)、およびIGSFP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、IGSFP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:1-12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:1-6およびSEQ ID NO:8-12からなる群から選択した或るアミノ酸配列と少なくとも90%が同一であるような天然アミノ酸配列を含むポリペプチド

(c)SEQ ID NO:1-12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

(d)SEQ ID NO:1-12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

10

【請求項 2】

SEQ ID NO:1-12からなる群から選択したアミノ酸配列を有する、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に記載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に記載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:13-24からなる群から選択したポリヌクレオチド配列を含む、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連結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 に記載のポリペプチドを生産する方法であって、

(a)前記ポリペプチドの発現に好適な条件下で、請求項 1 に記載された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連結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る細胞を培養する過程と、

30

(b)そのように発現した前記ポリペプチドを回収する過程とからなる方法。

【請求項 10】

前記ポリペプチドがSEQ ID NO:1-12からな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9に記載の方法。

【請求項 11】

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 12】

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

40

(a)SEQ ID NO:13-24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:13-18およびSEQ ID NO:20-24からなる群から選択した或るポリヌクレオチド配列に対して少なくとも90%が同一であるような天然ポリヌクレオチド配列を含むポリヌクレオチド

(c)SEQ ID NO:19のポリヌクレオチド配列に対して少なくとも94%が同一であるような天然ポリヌクレオチド配列を含むポリヌクレオチド

(d)(a)のポリヌクレオチドに相補的なポリヌクレオチド

(e)(b)のポリヌクレオチドに相補的なポリヌクレオチド

(f)(c)のポリヌクレオチドに相補的なポリヌクレオチド、および

50

(e)(a) ~ (f)のRNA等価物

【請求項13】

請求項12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14】

請求項12に記載のポリヌクレオチドの配列を有する標的ポリヌクレオチドをサンプル中から検出する方法であって、

(a)前記サンプル中の前記標的ポリヌクレオチドに相補的な配列を持つ少なくとも20の連続したヌクレオチドを含むプローブを用いて前記サンプルをハイブリダイズする過程であって、前記プローブと前記標的ポリヌクレオチドあるいはその断片との間でハイブリダイゼーション複合体が形成される条件下で、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過程と、

(b)前記ハイブリダイゼーション複合体の有無を検出し、該複合体が存在する場合にはオプションでその量を検出する過程を含む方法。

【請求項15】

前記プローブが少なくとも60の連続したヌクレオチドを含むことを特徴とする請求項14に記載の方法。

【請求項16】

請求項12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)前記の増幅した標的ポリヌクレオチドまたはその断片の有無を検出し、該標的ポリヌクレオチドまたはその断片が存在する場合にはオプションでその量を検出する過程を含むことを特徴とする方法。

【請求項17】

請求項1に記載のポリペプチドと、薬剤として許容できる賦形剤とを含む組成物。

【請求項18】

前記ポリペプチドが、SEQ ID NO:1 - 12からなる群から選択した或るアミノ酸配列を含むことを特徴とする、請求項17に記載の組成物。

【請求項19】

機能的なIGSFPPの発現の低下に関連する疾患や病態の治療方法であって、そのような治療が必要な患者に請求項17に記載の組成物を投与することを特徴とする治療方法。

【請求項20】

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

(a)請求項1に記載のポリペプチドを含むサンプルを化合物に曝すステップと、

(b)前記サンプルにおいてアゴニスト活性を検出する過程を含むことを特徴とする方法。

【請求項21】

請求項20に記載の方法によって同定したアゴニスト化合物と、薬剤として許容できる賦形剤とを含むことを特徴とする組成物。

【請求項22】

機能的なIGSFPPの発現の低下に関連する疾患や病態の治療方法であって、そのような治療が必要な患者に請求項21に記載の組成物を投与することを特徴とする治療方法。

【請求項23】

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であって、

(a)請求項1に記載のポリペプチドを含むサンプルを化合物に曝すステップと、

10

20

30

40

50

(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過程を含むことを特徴とする方法。

【請求項24】

請求項23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25】

機能的IGSFPの過剰発現に関連する疾患や病態の治療方法であって、そのような治療が必要な患者に請求項24に記載の組成物を投与することを含むことを特徴とする治療方法。

【請求項26】

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)請求項1に記載のポリペプチドを適切な条件下で少なくとも1つの試験化合物と混合する過程と、

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

【請求項27】

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

(a)請求項1のポリペプチドの活性が許容される条件下で、請求項1に記載のポリペプチドを少なくとも1つの試験化合物と混合する過程と、 20

(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28】

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

(a)前記標的ポリヌクレオチドの発現に好適な条件下で、該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、 30

(b)前記標的ポリヌクレオチドの発現改変を検出する過程と、

(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29】

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)処理した前記生物学的サンプルの核酸と、請求項12に記載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持つプローブをハイブリダイズさせる過程であって、このハイブリダイゼーションが、前記プローブと前記生物学的サンプル中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12に記載のポリヌクレオチドまたはその断片のポリヌクレオチド配列を持つポリヌクレオチドである、前記過程と、 40

(c)ハイブリダイゼーション複合体の収量を定量する過程と、

(d)前記処理された生物学的サンプル中の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ような方法。

【請求項30】

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

(a)前記生物学的サンプルと請求項11に記載の抗体との混合を、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を形成するのに適した条件下で行う過程と、

(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31】

請求項11に記載の抗体が、

(a)キメラ抗体

(b)単鎖抗体

(c)Fab断片

(d) $F(ab')_2$ 断片、あるいは

(e)ヒト化抗体のいずれかである抗体。

【請求項32】

請求項11に記載の抗体と、許容できる賦形剤とを含む組成物。

【請求項33】

被検者におけるIGSFPの発現に関連する症状又は疾患の診断方法であって、請求項32に記載の組成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34】

前記抗体が標識されることを特徴とする請求項32に記載の組成物。

【請求項35】

被検者におけるIGSFPの発現に関連する症状又は疾患の診断方法であって、請求項34に記載の組成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36】

請求項11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:1-12からなる群から選択したアミノ酸配列またはその免疫原性断片を含むポリペプチド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前記ポリペプチドでスクリーニングし、それによって、SEQ ID NO:1-12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に特異的に結合するポリクローナル抗体を同定する過程とを含むような方法。

【請求項37】

請求項36に記載の方法で産出したポリクローナル抗体。

【請求項38】

請求項37に記載のポリクローナル抗体及び適切なキャリアを含む組成物。

【請求項39】

請求項11に記載の抗体の特異性を有するモノクローナル抗体を作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:1-12からなる群から選択したアミノ酸配列またはその免疫原性断片を含むポリペプチド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)前記抗体産出細胞と不死化した細胞とを融合して、モノクローナル抗体を産出するハイブリドーマ細胞を形成する過程と、

(d)前記ハイブリドーマ細胞を培養する過程と、

(e)SEQ ID NO:1-12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に特異的に結合するようなモノクローナル抗体を前記培養物から単離する過程とを含むことを特徴とする方法。

【請求項40】

請求項39に記載の方法で産出したモノクローナル抗体。

【請求項41】

請求項40に記載のモノクローナル抗体及び適切なキャリアを含む組成物。

【請求項42】

10

20

30

40

50

Fab発現ライブラリのスクリーニングにより前記抗体を産出することを特徴とする請求項11に記載の抗体。

【請求項43】

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11に記載の抗体。

【請求項44】

SEQ ID NO:1-12を含む群から選択したアミノ酸配列を有するポリペプチドをサンプル中で検出する方法であって、

(a)請求項11に記載の抗体と前記ポリペプチドとの特異結合を許容する条件下で、前記抗体と1サンプルとをインキュベートする過程と、

(b)特異結合を検出する過程とを含み、該特異結合が、SEQ ID NO:1-12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がサンプル中に存在することを標示することを特徴とする方法。

10

【請求項45】

SEQ ID NO:1-12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をサンプルから精製する方法であって、

(a)請求項11に記載の抗体と前記ポリペプチドとの特異結合を許容する条件下で、前記抗体と1サンプルとをインキュベートする過程と、

(b)前記サンプルから前記抗体を分離し、SEQ ID NO:1-12からなる群から選択したアミノ酸配列を含む精製したポリペプチドを得る過程とを含むことを特徴とする方法。

20

【請求項46】

マイクロアレイの少なくとも1つのエレメントが請求項13に記載のポリヌクレオチドであるマイクロアレイ。

【請求項47】

ポリヌクレオチドを含むサンプルの発現プロフィールを作製する方法であって、

(a)サンプル中のポリヌクレオチドを標識化する過程

(b)ハイブリダイゼーション複合体が形成されるのに適した条件下で請求項46に記載のマイクロアレイのエレメントとサンプル中の標識化ポリヌクレオチドを接触させる過程

(c)サンプル中のポリヌクレオチドの発現を定量する過程を含む方法

【請求項48】

固体基板上の別個の物理的位置に固定した異なるヌクレオチド分子群を含むアレイであり、該ヌクレオチド分子の少なくとも1つが標的ポリヌクレオチドの少なくとも30の連続するヌクレオチドと特異的にハイブリダイゼーション可能な最初のオリゴヌクレオチドまたはポリヌクレオチド配列を含み、前記標的ポリヌクレオチドが請求項12に記載のポリヌクレオチドであるアレイ。

30

【請求項49】

請求項48に記載のアレイであり、前記最初のオリゴヌクレオチドまたはポリヌクレオチド配列は前記標的ポリヌクレオチドの少なくとも30の連続するヌクレオチドに完全に相補的であるアレイ。

【請求項50】

請求項48に記載のアレイであり、前記最初のオリゴヌクレオチドまたはポリヌクレオチド配列は前記標的ポリヌクレオチドの少なくとも60の連続するヌクレオチドに完全に相補的であるアレイ。

40

【請求項51】

請求項48に記載のアレイであり、前記最初のオリゴヌクレオチドまたはポリヌクレオチド配列が前記標的ポリヌクレオチドに完全に相補的であるアレイ。

【請求項52】

請求項48に記載のアレイであり、マイクロアレイであるアレイ。

【請求項53】

請求項48に記載のアレイであり、前記最初のオリゴヌクレオチドまたはポリヌクレオチド

50

配列を含むヌクレオチド分子にハイブリダイズされた前記標的ポリヌクレオチドをも含むアレイ。

【請求項 5 4】

請求項48に記載のアレイであり、リンカーが少なくとも1つの前記ヌクレオチド分子を前記固体基板に接合するアレイ。

【請求項 5 5】

請求項48に記載のアレイであり、基板上のそれぞれ固有の物理的位置には複数のヌクレオチド分子を含み、またいかなる単一の固有の物理位置の該複数のヌクレオチド分子も同じ配列を有し、基板上の各固有物理的位置には基板上の別の固有の物理的位置でのヌクレオチド分子の配列と異なる配列を持つヌクレオチド分子が含まれるアレイ。

10

【請求項 5 6】

SEQ ID NO:1のアミノ酸配列を含む請求項1に記載のポリペプチド。

【請求項 5 7】

SEQ ID NO:2のアミノ酸配列を含む請求項1に記載のポリペプチド。

【請求項 5 8】

SEQ ID NO:3のアミノ酸配列を含む請求項1に記載のポリペプチド。

【請求項 5 9】

SEQ ID NO:4のアミノ酸配列を含む請求項1に記載のポリペプチド。

【請求項 6 0】

SEQ ID NO:5のアミノ酸配列を含む請求項1に記載のポリペプチド。

20

【請求項 6 1】

SEQ ID NO:6のアミノ酸配列を含む請求項1に記載のポリペプチド。

【請求項 6 2】

SEQ ID NO:7のアミノ酸配列を含む請求項1に記載のポリペプチド。

【請求項 6 3】

SEQ ID NO:8のアミノ酸配列を含む請求項1に記載のポリペプチド。

【請求項 6 4】

SEQ ID NO:9のアミノ酸配列を含む請求項1に記載のポリペプチド。

【請求項 6 5】

SEQ ID NO:10のアミノ酸配列を含む請求項1に記載のポリペプチド。

30

【請求項 6 6】

SEQ ID NO:11のアミノ酸配列を含む請求項1に記載のポリペプチド。

【請求項 6 7】

SEQ ID NO:12のアミノ酸配列を含む請求項1に記載のポリペプチド。

【請求項 6 8】

SEQ ID NO:13のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

【請求項 6 9】

SEQ ID NO:14のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

【請求項 7 0】

SEQ ID NO:15のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

40

【請求項 7 1】

SEQ ID NO:16のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

【請求項 7 2】

SEQ ID NO:17のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

【請求項 7 3】

SEQ ID NO:18のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

【請求項 7 4】

SEQ ID NO:19のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

【請求項 7 5】

SEQ ID NO:20のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

50

【請求項 76】

SEQ ID NO:21のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

【請求項 77】

SEQ ID NO:22のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

【請求項 78】

SEQ ID NO:23のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

【請求項 79】

SEQ ID NO:24のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

【発明の詳細な説明】

【技術分野】

10

【0001】

本発明は、免疫グロブリンスーパーファミリータンパク質の核酸配列及びアミノ酸配列に関する。本発明はまた、これらの配列を利用した免疫系の疾患、神経の疾患、発生または発達障害、筋疾患および細胞増殖異常の診断・治療・予防に関する。本発明はさらに、免疫グロブリンスーパーファミリータンパク質の核酸配列及びアミノ酸配列の発現における外来性化合物の影響についての評価に関する。

【背景技術】

【0002】

認識、接着または結合などの機能を仲介する多くの細胞表面分子および可溶性分子は、進化における1つの共通の前駆体から進化したものである（すなわち、これらのタンパク質は構造的類似性を示す）。同様の機能を持つ免疫系以外の数多くの分子も、この同じ進化上の前駆体由来する。これらの分子は、免疫グロブリン(Ig)スーパーファミリーのメンバーとして分類化される。タンパク質がIgスーパーファミリーのメンバーであるための基準は、1つ以上のIgドメインを持ち、それが70～110アミノ酸残基の長さの領域であり、Ig可変領域様(V)ドメインまたはIg定常領域様(C)ドメインと相同であることである。Igスーパーファミリーのメンバーには抗体(Ab)、T細胞受容体(TCR)、クラスIおよびIIの主要組織適合性(MHC)タンパク質、CD2、CD3、CD4、CD8、ポリIg受容体、Fc受容体、神経細胞接着分子(NCAM)および血小板由来成長因子受容体(PDGFR)などが含まれる。

20

【0003】

Igドメイン(VおよびC)は保存されたアミノ酸残基の領域であり、ポリペプチドに、免疫グロブリン(または抗体)フォールドと呼ばれる球状3次構造を与える。これは、ほぼ平行なシートの2層からなる。保存的なシステイン残基が、55～75アミノ酸残基の長さをもつ、鎖内ジスルフィド結合ループを形成し、これがこれらシートの2層を接続する。各シートは、3または4本の逆平行ストランドを持ち、各ストランドは5～10アミノ酸残基の長さである。ストランド群内のアミノ酸残基の疎水性および親水性の相互作用がIgフォールドを安定させる(疎水性領域はストランドの内向きに面しているアミノ酸残基にあり、親水性領域は外向きに面している部分のアミノ酸残基にある)。VドメインはCドメインより長いポリペプチドで構成され、Igフォールド内に、付加的な1対のストランドを持つ。

30

40

【0004】

Igスーパーファミリー遺伝子の一貫した特徴は、あるIgドメインの各配列が1つのエキソンによってコードされることである。Igスーパーファミリーは、細胞間相互作用の仲介に関与する1つのIgドメインをコードする遺伝子から進化した可能性がある。そしてスーパーファミリーの新たなメンバーは、エキソンおよび遺伝子の複製によって生じた。現代のIgスーパーファミリータンパク質は、異なる数のVドメインおよび/またはCドメインを有する。このスーパーファミリーの進化上の別の特徴は、DNAの再編成を起こす能力である。これは独自の機能で、ファミリーの抗原受容体メンバーが保持している。

【0005】

Igスーパーファミリーのメンバーの多くは膜内在性の形質膜タンパク質であり、細胞外Ig

50

ドメインを持つ。それらの膜貫通ドメインおよびそれらの細胞質内尾部の疎水性アミノ酸残基は非常に多様で、Igファミリーのメンバー内または既知のシグナル伝達をする構造との相同性は少ないかまたは欠いている。スーパーファミリーに関するこの一般的な記述には例外がいくつかある。例えば、PDGFRの細胞質内尾部は、チロシンキナーゼ活性を持つ。また、Thy-1は、胸腺細胞およびT細胞で見られる糖タンパク質である。このタンパク質は細胞質内尾部を持たないが、その代わりに原形質膜に共有結合によるグリコシルホスファチジルイノシトール結合で固着している。

【0006】

Igスーパーファミリータンパク質の多くはまた、これらの分子の機能に必須の、Igドメイン間の相互作用という別の共通の特徴を持っている。多量体タンパク質のIgドメイン間の相互作用には、同種親和性のものと異種親和性（すなわち、同じIgドメイン間の作用、または異なるIgドメイン間の作用）のものがある。抗体は多量体タンパク質であり、Igドメイン間の同種親和性と異種親和性の双方の相互作用を持つ。重鎖の定常領域の対が抗体のFc領域を形成し、軽鎖および重鎖の可変領域の対が抗体の抗原結合部位を形成する。異種親和性の相互作用はまた、異なる分子のIgドメイン間でも起こる。これらの相互作用は免疫系における、または発生中および成熟した神経系における重要な細胞間相互作用のための細胞間の接着を提供する。（Abbas, A.K. 他（1991）Cellular and Molecular Immunology, W.B. Saunders Company, Philadelphia, PA, 142-145ページの概説を参照）。

10

【0007】

抗体

抗体はIgスーパーファミリーの多量体メンバーであり、B細胞の表面で発現されるか、または、B細胞によって分泌されて血液循環に入る。抗体は、血液中や他の細胞外液中で外来抗原に結合し、それらを中和する。プロトタイプの抗体は、ジスルフィド結合によって連結された2つの同じポリペプチド重鎖（H鎖）と2つの同じポリペプチド軽鎖（L鎖）からなる四量体である。この配列は、抗体分子に対して特徴的なY型を形成する。抗体はH鎖組成に基づいて分類される。抗体の5つのクラスであるIgA、IgD、IgE、IgGおよびIgMは、 α 、 δ 、 ϵ 、 γ および μ のH鎖型によって定義される。L鎖には κ と λ の2つのタイプがあり、どちらも、対としていずれかのH鎖対に会合し得る。血液循環内に見られる抗体の最も一般的なクラスであるIgGは四量体であるが、抗体の他のクラスのものは一般にこの基本的構造の変異体か、または多量体である。

20

30

【0008】

H鎖とL鎖は各々、N末端可変領域とC末端定常領域を有する。定常領域はL鎖の約110のアミノ酸と、H鎖の約330または440のアミノ酸から構成される。定常領域のアミノ酸配列は、或る特定クラスのH鎖またはL鎖群の内では、ほぼ同一である。可変領域は約110のアミノ酸からなり、H鎖とL鎖の両方にある。しかし、可変領域のアミノ酸配列は、特定クラスのH鎖またはL鎖群の中でも異なる。H鎖またはL鎖の可変領域のそれぞれに、広範な配列多様性を持つ3つの高頻度可変領域があり、各々約5~10のアミノ酸からなる。抗体分子において、H鎖およびL鎖の高頻度可変領域は1つになり、抗原認識部位を形成する（Alberts, B. 他（1994）Molecular Biology of the Cell, Garland Publishing, New York, NY, 1206-1213及び1216-1217ページに概説）。

40

【0009】

H鎖とL鎖は共に、Igスーパーファミリーのメンバーの、反復したIgドメインを含む。例えば、典型的なH鎖は4つのIgドメインを含んでおり、そのうちの3つは定常領域内に発生し、1つは可変領域内に発生して抗原認識部位の形成に寄与している。同様にして、典型的なL鎖は2つのIgドメインを含み、そのうちの1つは定常領域内に発生し、他の1つは可変領域内に発生する。

【0010】

免疫系は体内に入る外来分子を認識し、それに応答する能力を持っている。したがって、免疫系はすべての可能性のある抗原に対する抗体の完全な蓄積で武装していなければならない。このような抗体の多様性は、可変領域と定常領域とをコードする遺伝子セグメント

50

群の体細胞性再配列によって作られる。これらの遺伝子セグメント群は、各々の遺伝子セグメントが隣接する高度に保存されたDNA配列間で生じる、部位特異的組換えによって連結されている。何百という異なった遺伝子セグメントがあるため、何百万という独自の遺伝子が、組み合わせにより作成され得る。その上、これらのセグメントの不正確な連結とこれらのセグメント内での異常に高頻度の体細胞性突然変異が、多様な抗体集団の産生に更に寄与している。

【0011】

神経細胞接着タンパク質

神経細胞接着タンパク質 (NCAP) は、神経系の発生および再生中の神経ネットワークの確立に役割を果たしている (Uyemura, K. 他 (1996) *Essays Biochem.* 31:37-48、BrummendorfおよびRathjen (1996) *Curr. Opin. Neurobiol.* 6:584-593)。NCAPは神経細胞の遊走、細胞接着、突起進展、軸索の繊維束形成、経路発見、シナプス認識、シナプス形成、髄鞘形成および再生に関与する。NCAPは学習と記憶に関連するニューロンの表面で発現されている。NCAPSをコードする遺伝子の突然変異は、遺伝性の神経障害であるCharcotMarieTooth病、DejerineSottas病、X遺伝子連鎖の水頭症、MASA症候群 (精神薄弱、失語症、引きずり歩行、および親指の内転) およびタイプIの痙性対麻痺等の神経疾患に関連付けられている。NCAPの発現は神経系に限られていない場合もある。たとえば、L1は黒色腫細胞および造血腫瘍細胞で発現していて、細胞の進展と遊走に関与しているとされ、腫瘍の進行に役割を果たしている可能性がある (Montgomery 他 (1996) *J. Cell Biol.* 132:475-485)。

10

20

【0012】

NCAPは少なくとも一つの免疫グロブリンの定常ドメインまたは可変ドメインを有する (Uyemura 他、前出)。これらは一般に膜貫通型ドメインおよび/またはグリコシルホスファチジルイノシトール (GPI) アンカーを通じて原形質膜に結合されている。GPI結合はGPIホスホリパーゼCによって切断されうる。ほとんどのNCAPは、1つ以上の免疫グロブリンドメインからなる細胞外領域、1つの膜貫通ドメイン、および1つの細胞内領域からなる。多くのNCAPは、共有結合によって付いたオリゴ糖、グルクロン酸、および硫酸を含む翻訳後修飾を有する。NCAPは、シンプルタイプ、コンプレックスタイプおよびミックスタイプという3つのサブグループに分かれる。シンプルタイプNCAPは一つ以上の可変免疫グロブリンドメインまたは定常免疫グロブリンドメインを含むが、その他のドメインを欠いている。シンプルタイプサブグループのメンバーとしては、シュワン細胞ミエリンタンパク質 (SMP)、辺縁系関連膜タンパク質 (LAMP)、アヘン剤結合細胞接着分子 (OBCAM) 等が挙げられる。コンプレックスタイプNCAPは免疫グロブリンドメインに加えてフィブロネクチンタイプIIIドメインを含んでいる。コンプレックスタイプサブグループは神経細胞接着分子 (NCAM)、axonin-1、F11、BravoおよびL1を含んでいる。ミックスタイプNCAPは、免疫グロブリンドメインと他のモチーフ (チロシンキナーゼ、上皮成長因子様ドメイン、セマドメイン、PSI (プレキシン、セマフォリン、インテグリン) ドメイン) を組み合わせて含んでいる。このサブグループは、神経成長因子 (NGF) およびノイロトロピン4 (NT4) 等の神経成長因子のTrk受容体、グリア成長因子II (GGFII) およびアセチルコリン受容体誘導因子 (ARIA) 等のNeu分化因子、セマフォリンBおよびコラプシン等のセマフォリン/コラプシンファミリー、及びプレキシン等のセマフォリン/コラプシンファミリーのメンバーの受容体 (プレキシンについては下記を参照) を含んでいる。

30

40

【0013】

NCAPサブファミリーの一つであるNCAP-LONサブグループは、脳のニューロンの固有の小集団において発現される細胞接着タンパク質を含んでいる。NCAP-LONサブグループのメンバーは3つの免疫グロブリンドメインを持ち、GPIアンカーを通じて細胞膜に結合する。たとえば、kilon (NCAP-LONの近縁) は脳の大脳皮質と海馬で発現される (Funatsu 他 (1999) *J. Biol. Chem.* 274:8224-8230)。免疫染色によると、Kilonは垂体ニューロンの樹状突起と細胞体に局在している。Kilonは3つのC2タイプ免疫グロブリン様ドメイン、6つの予想されるグリコシル化部位および1つのGPIアンカーを持っている。Kilonの発現は

50

発生的に調節されている。胚および出産後初期の脳と比べて、成人の脳で発現レベルが高くなっている。共焦点顕微鏡により、Kilonは神経ペプチド、オキシトシンまたはアルギニンバソプレッシンを分泌する視床下部大細胞ニューロンの樹状突起に存在することが示されている(Miyata 他 (2000) *J. Comp. Neurol.* 424:74-85)。アルギニンバソプレッシンは体液のホメオスタシス、細胞外のモル浸透圧濃度および血管内容積を調節する。オキシトシンは出産時の子宮平滑筋および授乳時の乳腺の筋上皮細胞の収縮を誘発する。大細胞ニューロンにおいて、Kilonは神経ペプチド分泌時の樹状突起接続の再構成に役割を果たしていると提唱されている。

【0014】

サイドキック(SDK)はNCAPファミリーのメンバーである。SDKの細胞外領域は、6つの免疫グロブリンドメインと13のフィブロネクチンタイプIIIドメインを含んでいる。SDKはショウジョウバエの眼の発達中に細胞-細胞相互作用に参与する(Nguyen, D. N. T. 他 (1997) *Development* 124: 3303)。

【0015】

シナプス膜糖タンパク質

特殊な細胞接着が、細胞-細胞接触のポイントで起こり得る。これらの細胞接着の中には、細胞間の化学及び電気信号の通過を仲介する連絡結合がある。中枢神経系では、神経間の連絡結合はシナプス結合として知られる。それらはシナプス前及びシナプス後ニューロンの膜と細胞骨格から成る。シナプス膜(SM)及び後シナプス密度肥厚(PSD)分画等の生化学的に単離したシナプス亜分画に見出される幾つかの糖タンパク質が同定され、また機能が確立されてきた。1つの事例は、 Na^+/K^+ -ATPaseの2サブユニットとして同定されるSM糖タンパク質であるgp50である。

【0016】

2つの糖タンパク質であるgp65とgp55は、ラットの前脳から作製したシナプス膜の主要な成分である。それらは、それぞれ3つ及び2つのIgドメインを含むIgスーパーファミリーのメンバーのメンバーである。Igスーパーファミリーのメンバーとして、これらのタンパク質の可能性のある機能がシナプス結合における接着相互作用を仲介するためのものと提起されている(Langnaese, K. 他 (1997) *J. Biol. Chem.* 272:821-827)。

【0017】

レクチン

レクチンは、細胞表面の糖質に特異的かつ可逆的に結合し、細胞の凝集をもたらす普遍的な細胞外糖タンパク質ファミリーである(Drickamer, K.およびTaylor, M. E. (1993) *Annu. Rev. Cell Biol.* 9:237-264の概説を参照)。この機能は、免疫反応の活性化に特に重要である。レクチンは炎症部位におけるリンパ球の凝集と分裂誘起刺激を仲介する(Lasky, L. A. (1991) *J. Cell. Biochem.* 45:139-146、Paietta, E. 他 (1989) *J. Immunol.* 143:2850-2857)。

【0018】

シアリン酸結合Ig様レクチン(SIGLEC)は、糖タンパク質と糖脂質のシアリン酸に結合するIgスーパーファミリーのメンバーである。SIGLECには、シアロアドヘシン(sialoadhesin)、CD22、CD33、ミエリン結合性糖タンパク質(MAG)、SIGLEC-5、SIGLEC-6、SIGLEC-7、SIGLEC-8が含まれる。SIGLECの細胞外領域は、膜末端V-セットドメインを有し、次いで異なる数のC2-セットドメインがその後続く。シアリン酸結合ドメインは、V-セットドメインにマッピングされる。神経系だけで発現するMAGを除いて、ほとんどのSIGLECは造血細胞の固有のサブセット上で発現する。例えばSIGLEC-8は、多形核白血球(顆粒球)の1つの形態である好酸球だけで発現する(Floyd, H.他 (2000) *J. Biol. Chem.* 275: 861-866)。

【0019】

ロイシンリッチリピートタンパク質

ロイシンリッチリピートタンパク質(LRRP)は、タンパク質-タンパク質相互作用に参与する。哺乳動物神経のロイシンリッチリピートタンパク質(NLLR-1及びNLLR-2)、ショウジョ

ウバエのコネクチン、slit、chaopin、toll等のLRRPは全て神経の発達に關与する。LRRPの細胞外領域には、様々な数のロイシンリッチリピート、免疫グロブリン様ドメイン、フィブロネクチンタイプIIIドメインが含まれる(Taguchi, A. 他 (1996) Brain Res. Mol. Brain Res. 35:31-40)。

【0020】

免疫グロブリン様ドメインのV及びC2セットに加えて、IPT/TIG (immunoglobulin-like fold shared by plexins and transcription factorsの略) と名付けられたDセット免疫グロブリン様ドメインがある。IPT/TIG含有タンパク質は、プレキシン、MET/ RON/ SEA (肝細胞成長因子受容体ファミリー)、アフリカツメガエルの1次神経細胞の特異化に關与するCol/Olf1/EBファミリーの転写因子である転写因子XCoe2を含む(Bork, P. 他. (1999) Trends in Biochem. 24:261-263、Santoro, N. M. 他. (1996) Mol. Cell Biol. 16:7072-7083、Dubois L. 他 (1998) Curr. Biol. 8:199209)。プレキシンA等のプレキシンとVESPRは、軸索誘導を制御するニューロンセマフォリン受容体であることが示されている(Winberg M. L. 他. (1998) Cell 95:903916)。

10

【0021】

発現プロファイル作成

アレイ技術は、単一の多型遺伝子の発現や、多数の関連遺伝子または無関係の遺伝子の発現プロファイルを探求する、簡単な方法を提供し得る。単一遺伝子の発現を試験するときには、アレイを用いて、或る特定遺伝子又はその変異体の発現を検出する。発現プロファイルを試験するときには、アレイは次のような遺伝子を同定するプラットフォームを提供する。即ちどの遺伝子が組織特異的か、毒性アッセイにおいてテストされる物質に影響されるか、シグナル伝達カスケードの一部であるか、ハウスキーピング機能を実行するか、又は、特定の遺伝的素因や、条件、疾患、又は障害に、特異的に関連する遺伝子であるかの同定である。

20

【0022】

新規の免疫グロブリンスーパーファミリータンパク質、およびそれらをコードするポリヌクレオチドの発見により、新規の組成物を提供することで当分野の要望に応えることができる。この新規の組成物は、免疫系の疾患、神経の疾患、発生または発達障害、筋疾患および細胞増殖異常の診断・予防・治療において有用であり、また、免疫グロブリンスーパーファミリータンパク質の核酸配列及びアミノ酸配列の発現における外来性化合物の影響

30

【発明の開示】

【発明の効果】

【0023】

本発明は、総称して「IGSFP」、個別にはそれぞれ「IGSFP-1」、「IGSFP-2」、「IGSFP-3」、「IGSFP-4」、「IGSFP-5」、「IGSFP-6」、「IGSFP-7」、「IGSFP-8」、「IGSFP-9」、「IGSFP-10」、「IGSFP-11」、および「IGSFP-12」と呼ぶ免疫グロブリンスーパーファミリータンパク質である精製されたポリペプチドを提供する。或る態様において本発明は、(a) SEQ ID NO:1-12からな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:1-12からなる群から選択したアミノ酸配列と少なくとも90%が同一である天然のアミノ酸配列からなるポリペプチド、(c) SEQ ID NO:1-12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:1-12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なる群から選択した単離されたポリペプチドを提供する。一態様では、SEQ ID NO:1-12のアミノ酸配列からなる単離されたポリペプチドを提供する。

40

【0024】

さらに、本発明は(a) SEQ ID NO:1-12からなる群から選択した或るアミノ酸配列からなるポリペプチド、(b) SEQ ID NO:1-12からなる群から選択した或るアミノ酸配列との少なくとも90%の同一性を持つ或る天然アミノ酸配列からなるポリペプチド、(c) SEQ ID NO:1-12からなる群から選択した或るアミノ酸配列を持つポリペプチドの生物学的活性断

50

片、または (d) SEQ ID NO:1-12からなる群から選択した或るアミノ酸配列を持つポリヌクレオチドの免疫原性断片、からなる群から選択した或るポリペプチドをコードする、単離されたポリヌクレオチドを提供する。一実施態様では、該ポリヌクレオチドは、SEQ ID NO:1-12からなる群から選択した或るポリペプチドをコードする。別の実施態様では、ポリヌクレオチドはSEQ ID NO:13-24からなる群から選択される。

【0025】

本発明は更に、(a) SEQ ID NO:1-12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:1-12を有する群から選択したアミノ酸配列と少なくとも90%が同一である天然のアミノ酸配列からなるポリペプチド、(c) SEQ ID NO:1-12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:1-12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から構成される群から選択されたポリペプチドをコードするポリヌクレオチドと機能的に連結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

【0026】

また、本発明は、(a) SEQ ID NO:1-12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:1-12からなる群から選択したアミノ酸配列と少なくとも90%が同一である天然のアミノ酸配列を有するポリペプチド、(c) SEQ ID NO:1-12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:1-12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し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回収する過程とを有し、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連結したプロモーター配列を有する。本発明は更に、(a) SEQ ID NO:1-12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:1-12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:1-12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:1-12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単離された抗体を提供する。

20

30

【0027】

本発明は更に、(a) SEQ ID NO:13-24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:13-24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) のポリヌクレオチドに相補的なポリヌクレオチド、(d) (b) のポリヌクレオチドに相補的なポリヌクレオチド、および (e) (a) ~ (d) のRNA等価物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

【0028】

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:13-24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:13-24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) のポリヌクレオチドに相補的なポリヌクレオチド、(d) (b) のポリヌクレオチドに相補的なポリヌクレオチド、および (e) (a) ~ (d) のRNA等価物からなる群から選択されたポリヌクレオチドの配列を有する。検出方法は、(a) サンプル中の上記標的ポリヌクレオチドに相補的な或る配列からなる少なくとも20の連続したヌクレオチド群からなる或るプローブを用いて該サンプルをハイブリダイズ

50

する過程と、(b)該ハイブリダイゼーション複合体の有無を検出し、複合体が存在すればオプションでその量を検出する過程からなる。該プローブと該標的ポリヌクレオチドあるいはその断片との間でハイブリダイゼーション複合体が形成されるような条件下で、プローブは、該標的ポリヌクレオチドに対し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。

【0029】

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:13-24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:13-24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)のポリヌクレオチドに相補的なポリヌクレオチド、(d)(b)のポリヌクレオチドに相補的なポリヌクレオチド、および(e)(a)~(d)のRNA等価物からなる群から選択された配列のポリヌクレオチドを有する。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)増幅した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

【0030】

本発明は更に、有効量のポリペプチドと薬剤として許容できる賦形剤とを含む組成物を提供する。有効量のポリペプチドは、(a)SEQ ID NO:1-12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:1-12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:1-12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の生物学的活性断片、および(d)SEQ ID NO:1-12からなる群から選択したアミノ酸配列の免疫原性断片からなる群から選択される。一実施例では、SEQ ID NO:1-12からなる一群から選択されたアミノ酸配列を含む組成物を提供する。更に、本発明は、患者にこの組成物を投与することを含む、機能的IGSF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

【0031】

本発明はまた、(a)SEQ ID NO:1-12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:1-12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:1-12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、および(d)SEQ ID NO:1-12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なる群から選択されたポリペプチドのアゴニストとしての有効性を確認するために化合物をスクリーニングする方法を提供する。スクリーニング方法は、(a)ポリペプチドを有するサンプルを化合物に曝す過程と、(b)サンプル中のアゴニスト活性を検出する過程とを含む。別法では、本発明は、この方法によって同定されたアゴニスト化合物と好適な医薬用賦形剤とを含む組成物を提供する。更なる別法では、本発明は、機能的IGSF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し、その内にはそのような治療を必要とする患者にこの組成物を投与することが含まれる。

30

40

【0032】

本発明は更に、(a)SEQ ID NO:1-12からなる群から選択した或るアミノ酸配列を含むポリペプチド、(b)SEQ ID NO:1-12からなる群から選択した或るアミノ酸配列との少なくとも90%の同一性を有する天然アミノ酸配列を含むポリペプチド、(c)SEQ ID NO:1-12からなる群から選択したアミノ酸配列を持つポリペプチドの生物学的活性断片、および(d)SEQ ID NO:1-12からなる群から選択したアミノ酸配列を持つポリペプチドの免疫原性断片、からなる群から選択したポリペプチドのアンタゴニストとしての有効性につき、或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)該ポリペプチドを含むサンプルを或る化合物に曝す過程と、(b)サンプル中のアンタゴニスト活性を検出する過程からなる。別法で本発明は、この方法によって同定したアンタゴニスト

50

化合物と薬剤として許容できる賦形剤とを含む組成物を提供する。更なる別法では、本発明は、機能的IGSFPの過剰な発現に関連した疾患やその症状の治療方法を提供し、その内にはそのような治療を必要とする患者にこの組成物を投与することが含まれる。

【0033】

本発明は更に、(a) SEQ ID NO:1-12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:1-12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:1-12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、または(d) SEQ ID NO:1-12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を含む群から選択されたポリペプチドに特異結合する化合物をスクリーニングする方法を提供する。スクリーニング方法は、(a) ポリペプチドを適切な条件下で少なくとも1つの試験化合物と混合させる過程と、(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

【0034】

本発明は更に、(a) SEQ ID NO:1-12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:1-12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:1-12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、または(d) SEQ ID NO:1-12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を含む群から選択されたポリペプチドの活性を調節する化合物をスクリーニングする方法を提供する。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

【0035】

更に本発明は、SEQ ID NO:13-24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或る化合物に曝露する過程と、(b) この標的ポリヌクレオチドの発現の変化を検出する過程と、(c) 可変量の該化合物の存在下での標的ポリヌクレオチドの発現を該化合物の不存在下での発現と比較する過程を含む、該スクリーニング方法を提供する。

30

【0036】

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:13-24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:13-24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:13-24からなる群から選択したポリヌクレオチド配列を有するポリヌクレオチド、(ii) SEQ ID NO:13-24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%が同一である天然のポリヌクレオチド配列を有するポリヌクレオチド、(iii) (i) のポリヌクレオチドに相補的なポリヌクレオチド、(iv) (ii) のポリヌクレオチドに相補的なポリヌクレオチド、(v) (i) ~ (iv) のRNA等価物を含む群から選択する。或いは、標的ポリヌクレオチドは、上記(i) ~ (

40

50

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を有する。毒性の算定方法には更に(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。

【発明を実施するための最良の形態】

【0037】

(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ものではないことも併せて理解されたい。

【0038】

請求の範囲及び明細書中で用いている単数形の「或る」及び「その(この)」の表記は、文脈から明らかにそうでないとされる場合を除いて複数のものを指す場合もあることに注意しなければならない。従って、例えば「或る宿主細胞」は複数の宿主細胞を含み、その「抗体」は複数の抗体が含まれ、当業者には周知の等価物なども含まれる。

【0039】

本明細書中で用いる全ての技術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して用い得る細胞株、プロトコル、試薬及びベクターについて説明及び開示する目的で引用しているものである。本明細書のいかなる開示内容も、本発明が先行技術の効力によってこのような開示に対して先行する権利を与えられていないことを認めるものではない。

【0040】

(定義)

用語「IGSFP」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IGSFPのアミノ酸配列を指す。

【0041】

「アゴニスト」という用語は、IGSFPの生物活性を強化あるいは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組成物を挙げることができるが、これらはIGSFPと直接相互作用することによって、或いはIGSFPが関与する生物学的経路の構成成分に作用することによって、IGSFPの活性を調節する。

【0042】

「対立遺伝子変異体」は、IGSFPをコードする遺伝子の別の形態である。対立遺伝子変異体は、核酸配列における少なくとも1つの突然変異から作製し得る。また、変異RNAまたはポリペプチドを作製し得る。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。或る遺伝子は、その天然型の対立遺伝子変異配列を全く持たない場合もあり、1個以上持つこともある。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一般に、ヌクレオチドの自然な欠失、付加または置換に帰するものである。これら各変化は、単独で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。

【0043】

IGSFPをコードする「変容した/改変された」核酸配列に含む配列には、様々なヌクレオチドの欠失、挿入、或いは置換を生じた配列であって、IGSFPと同一のポリペプチドを、或いは、IGSFPの機能的特徴の少なくとも1つを備えるポリペプチドを生じる配列がある。この定義には、IGSFPをコードするポリヌクレオチド配列にとり正常な染色体の遺伝子

座ではない位置での、対立遺伝子変異配列群への不適當あるいは予期しないハイブリダイゼーションを含み、また、IGSFPをコードするポリヌクレオチドの特定のオリゴヌクレオチドプローブを用いて容易に検出可能な或いは検出困難な多型性を含む。コードされたタンパク質も「変異」し得るものであり、サイレント変化を生ぜしめて結果的に機能的に等価なIGSFPとなるようなアミノ酸残基の欠失、挿入または置換を含み得る。意図的なアミノ酸置換は、生物学的或いは免疫学的にIGSFPの活性が保持される範囲で、残基の、極性、電荷、溶解度、疎水性、親水性、および/または両親媒性についての、類似性に基いて成され得る。例えば、負に帯電したアミノ酸にはアスパラギン酸及びグルタミン酸があり、正に帯電したアミノ酸にはリジン及びアルギニンがある。親水性値が近似した非荷電極性側鎖を持つアミノ酸には、アスパラギンとグルタミン、セリンとトレオニンがある。親水性値が近似した非荷電側鎖を持つアミノ酸には、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

【0044】

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

【0045】

「増幅」は、核酸配列の複製物を作製することに関連する。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。

20

【0046】

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

【0047】

「抗体」の語は、抗原決定基と結合することができる、無傷の免疫グロブリン分子やその断片、例えばFab、F(ab')₂及びFv断片を指す。IGSFPポリペプチドと結合する抗体は、免疫抗原として、無傷ポリペプチド群を用いて、または、目的の小ペプチド群を有する断片群を用いて作製可能である。動物(マウス、ラットあるいはウサギ等)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、好みに応じてキャリアタンパク質に抱合することも可能である。通常用いられるキャリア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(KLH)等がある。その結合ペプチドを、動物を免疫化するために用いる。

30

【0048】

用語「抗原決定基」は、特定の抗体と接触する分子の領域(即ちエピトープ)を指す。タンパク質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、抗体への結合について、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。

40

【0049】

用語「アプタマー」は、特定の分子標的に結合する核酸またはオリゴヌクレオチド分子を指す。アプタマーはin vitroの進化過程(例えば米国特許番号第5,270,163号に記載されたSELEX(Systematic Evolution of Ligands by EXponential Enrichment))から由来するもので、そのような過程は大きな組み合わせライブラリから標的特異的なアプタマー配列を選択する。アプタマー組成は、2本鎖または1本鎖であってもよく、デオキシリボヌクレオチド、リボヌクレオチド、ヌクレオチド誘導體、または他のヌクレオチド様分子を含み得る。アプタマーのヌクレオチド成分は、修飾された糖基(例えばリボヌクレオチド

50

の2'-OH基が2'-Fまたは2'-NH₂で置換し得る)を有することが可能で、そのような糖基はヌクレアーゼへの抵抗性または血液中でのより長い寿命などの望ましい性質に改善し得る。循環系からアプタマーが除去される速度を遅くするために、アプタマーを高分子量キャリアー等の分子に抱合させることができる。アプタマーは、たとえば架橋剤の光活性化によって、各々の同種リガンドと特異的に架橋させることができる(Brody, E.N.及びL. Gold (2000) J. Biotechnol. 74:5-13等を参照)。

【0050】

「イントラマー(intramer)」の用語はin vivoで発現されるアプタマーを意味する。例えば、ワクシニアウイルスに基づく或るRNA発現系を用いて、白血球の細胞質内で特定のRNAアプタマー類が高レベルに発現されている(Blind, M.他(1999) Proc. Natl Acad. Sci. USA 96:3606-3610)。

【0051】

「スピーゲルマー(spiegelmer)」の語はL-DNA、L-RNAその他の左旋性ヌクレオチド誘導体またはヌクレオチド様分子を含むアプタマーを指す。左旋性のヌクレオチドを含むアプタマーは右旋性ヌクレオチドに作用する天然の酵素による分解に対して耐性がある。

【0052】

用語「アンチセンス」は、特定の核酸配列の「センス」(コーディング)鎖と塩基対を形成し得る任意の組成物を指す。アンチセンス組成物としては、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結合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を含みうる。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、細胞が産生した天然核酸配列との塩基対を形成し、転写または翻訳を阻止する二重鎖を形成する。「負」または「マイナス」という表現は、ある参考DNA分子のアンチセンス鎖を意味し、「正」または「プラス」という表現は、ある参考DNA分子のセンス鎖を意味しうる。

【0053】

用語「生物学的に活性」は、天然分子の構造的、調節的、或いは生化学的な機能を有するタンパク質を指す。同様に、用語「免疫学的に活性」または「免疫原性」は、天然或いは組換え体のIGSFP、合成のIGSFP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。

【0054】

「相補的」は、塩基対合によってアニールする2つの一本鎖核酸配列間の関係を指す。例えば、配列「5'A-G-T3'」は、相補配列「3'T-C-A5'」と対を形成する。

【0055】

「~のポリヌクレオチド配列を含む(有する)組成物」または「~のアミノ酸配列を含む(有する)組成物」は広い意味で、所定のポリヌクレオチド配列若しくはアミノ酸配列を含む任意の組成物を指す。この組成物には、乾燥製剤または水溶液が含まれ得る。IGSFP若しくはIGSFPの断片をコードするポリヌクレオチド配列を含む組成物は、ハイブリダイゼーションプローブとして使用され得る。このプローブは、凍結乾燥状態で保存可能であり、糖質などの安定化剤と結合させることが可能である。ハイブリダイゼーションにおいては、塩(例えばNaCl)、界面活性剤(例えばドデシル硫酸ナトリウム; SDS)及びその他の構成成分(例えばデンハート液、脱脂粉乳、サケ精子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。

【0056】

「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、X-L-PCRキット(Applied Biosystems, Foster City CA)を用いて5'及び/または3'の方

向に伸長され、再度シーケンシングされた核酸配列、またはGELVIEW断片構築システム (GCG, Madison, WI) またはPhrap (University of Washington, Seattle WA) 等の断片構築用のコンピュータプログラムを用いて1つ或いはそれ以上のオーバーラップする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。

【0057】

「保存的なアミノ酸置換」は、元のタンパク質の特性を殆ど変えない置換を指す。即ち、置換によってそのタンパク質の構造や機能が大きくは変わらず、そのタンパク質の構造、特にその機能が保存される。下表は、タンパク質中で元のアミノ酸と置換され得るアミノ酸であり、保存的なアミノ酸置換と認められるアミノ酸を示している。

10

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

20

30

【0058】

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 螺旋構造、(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。

【0059】

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

【0060】

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

40

【0061】

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

50

【0062】

「差次的発現」は、少なくとも2つの異なったサンプルを比較することによって決められる、増加（上方調節）、あるいは減少（下方調節）、または遺伝子発現の欠損またはタンパク発現の欠損を指す。このような比較は例えば、処理済サンプルと不処理サンプル、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。

【0063】

「エキソンシャフリング」は、異なるコード領域（エキソン）の組換えを意味する。1つのエキソンはコードされるタンパク質の1つの構造的または機能的ドメインを代表し得るため、安定した基礎構造の新規な再構築(reassortment)を介して新しいタンパク質が組立てられることが可能であり、これにより新しいタンパク質機能の進化を促進できる。

10

【0064】

用語「断片」は、IGSFPまたはIGSFP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、定義された配列の全長から1ヌクレオチド/アミノ酸残基を差し引いた長さよりも短い長さを有し得る。例えば或る断片は、5~1000の連続したヌクレオチドまたはアミノ酸残基を有し得る。プローブ、プライマー、抗原、治療用分子として、或いはその他の目的のために用いられる断片は、少なくとも5、10、15、16、20、25、30、40、50、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の長さであり得る。断片は、或る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ポリペプチドの最初の250または500アミノ酸（または最初の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

【0065】

SEQ ID NO:13-24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:13-24を特異的に同定するものであり、例えば断片が得られる同一ゲノム中の他すべての配列とは異なるものである。SEQ ID NO:13-24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:13-24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:13-24の断片の正確な長さ及び、断片が対応するSEQ ID NO:13-24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。

30

【0066】

SEQ ID NO:1-12のある断片は、SEQ ID NO:13-24のある断片によってコードされる。SEQ ID NO:1-12のある断片には、SEQ ID NO:1-12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:1-12のある断片は、SEQ ID NO:1-12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:1-12のある断片の正確な長さ、及びその断片に対応するSEQ ID NO:1-12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。

【0067】

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

40

【0068】

「相同性」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。

【0069】

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、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

50

しうるので、2つの配列をより有意に比較できる。

【0070】

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.および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を2つ1組でアラインメントする際のデフォルトパラメータは、Ktuple=2、gap penalty=5、window=4、「diagonals saved」=4と設定する。

「weighted(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る。

【0071】

一致率は、アラインメントされたポリヌクレオチド配列間の「percent similarity(類似性パーセント)」としてCLUSTAL Vによって報告される。

【0072】

或いは、一般的に用いられ且つ自由に入手できる配列比較アルゴリズム一式が、国立バイオテクノロジー情報センター(NCBI) Basic Local Alignment Search Tool(BLAST)から提供されており(Altschul, S.F. 他(1990) J. Mol. Biol. 215:403-410)、これは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/>)を含む幾つかの情報源から入手可能である。このBLASTソフトウェア一式には、既知のポリヌクレオチド配列と様々なデータベースの別のポリヌクレオチド配列とのアラインメントに用いられる「blastn」を含む、様々な配列分析プログラムが含まれる。「BLAST 2 Sequences」と呼ばれるツールも入手可能であり、2つのヌクレオチド配列を直接にペアワイズで比較するために用いられる。「BLAST 2 Sequences」は、<http://www.ncbi.nlm.nih.gov/gorf/bl2.html>にアクセスして、対話形式で利用ができる。「BLAST 2 Sequences」ツールは、blastn及びblastp(以下に記載)の両方に用いることができる。BLASTプログラムは、一般的には、ギャップ及び他のパラメータをデフォルト設定に設定して用いる。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。

【0073】

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

【0074】

一致率は、ある定義された配列の全長(例えば特定のSEQ IDナンバーで定義された配列)について測定し得る。或いは、より短い長さ、例えば、定義された、より大きな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)の長さについ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。

【0075】

高度の同一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

10

20

30

40

50

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。

【0076】

ポリペプチド配列に用いられる用語「一致率」または「%一致」とは、標準化されたアルゴリズムを用いてアラインメントされる2つ以上のポリペプチド配列間の一致する残基の百分率のことである。ポリペプチド配列アラインメントの方法は公知である。保存的アミノ酸置換を考慮するアラインメント方法もある。既に詳述したこのような保存的置換は通常、置換部位の酸性度及び疎水性を保存するので、ポリペプチドの構造を（従って機能も）保存する。

【0077】

ポリペプチド配列間の一致率は、MEGALIGN version 3.12e配列アラインメントプログラムに組み込まれているようなCLUSTAL Vアルゴリズムのデフォルトのパラメータを用いて決定できる（既に説明したのでそれを参照されたい）。CLUSTAL Vを用いて、ポリペプチド配列をペアワイズアラインメントする際のデフォルトパラメータは、Ktuple=1、gap penalty=3、window=5、「diagonals saved」=5と設定される。デフォルトの残基重み付け表としてPAM250マトリクスが選択される。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「percent similarity」として一致率を報告する。

10

【0078】

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ペアワイズで比較する場合、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 (2000年4月21日)でblastpを使用するであろう。デフォルトパラメータの設定例を以下に示す。

20

【0079】

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

30

【0080】

一致率は、ある定義されたポリペプチド配列の全長（例えば特定のSEQ IDナンバーで定義された配列）について測定し得る。或いは、より短い長さ、例えば、定義された、より大きな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）の長さについ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支持さ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。

【0081】

「ヒト人工染色体（HAC）」は、約6kb ~ 10MbのサイズのDNA配列を含み得る、安定した染色体複製の分離及び維持に必要な全てのエレメントを含む直鎖状の微小染色体である。用語「ヒト化抗体」は、もとの結合能力を保持しつつ、よりヒトの抗体に似せるために、非抗原結合領域のアミノ酸配列が変えられた抗体分子を指す。

40

【0082】

「ハイブリダイゼーション」とは、所定のハイブリダイゼーション条件下で、ある一本鎖ポリヌクレオチドがある相補的な一本鎖と塩基対を形成するアニーリングのプロセスである。特異的ハイブリダイゼーションは、2つの核酸配列が高い相補性を共有することの指標である。特異的ハイブリダイゼーション複合体は許容的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、よりストリ

50

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。核酸配列のアニールングのための許容的条件は、当業者が慣例的に決定できる。許容的条件は、どのハイブリダイゼーション実験でも一定でありうるが、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によって変更されうる。アニール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。

【0083】

一般に、ハイブリダイゼーションのストリンジェンシーは或る程度、洗浄ステップを実行する温度を基準にして表すことができる。このような洗浄温度は通常、所定のイオン強度及びpHにおける特異配列の融点（ T_m ）より約5～20 低くなるように選択する。この T_m は、所定のイオン強度及びpHの条件下で、完全に一致するプローブに標的配列の50%がハイブリダイズする温度である。 T_m を計算する式及び核酸のハイブリダイゼーション条件はよく知られており、Sambrook J 他（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。

10

【0084】

本発明の、ポリヌクレオチドとポリヌクレオチドのハイブリダイゼーションに対する高ストリンジェンシー条件には、約0.2×SSC及び約0.1%のSDS存在下で68 において1時間の洗浄条件が含まれる。或いは、65 、60 、55 または42 の温度で行ってもよい。SSC濃度は、約0.1%のSDS存在下で、約0.1～2×SSCの範囲で変化し得る。通常は、ブロッキング剤を用いて非特異ハイブリダイゼーションを阻止する。このようなブロッキング剤には、例えば、約100～200 μg/mlのせん断した変性サケ精子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、ヌクレオチド群及びヌクレオチドにコードされるポリペプチド群について、類似の役割を強く示唆している。

20

【0085】

用語「ハイブリダイゼーション複合体」は、相補的な塩基間の水素結合によって形成された、2つの核酸配列の複合体を指す。ハイブリダイゼーション複合体は、溶解状態で形成し得る（ C_0t または R_0t 解析等）。或いは、一方の核酸配列が溶解状態で存在し、もう一方の核酸配列が固体支持体（例えば紙、膜、フィルタ、チップ、ピンまたはガラススライド、或いは他の適切な基板であって細胞若しくはその核酸が固定される基板）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。

30

【0086】

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

【0087】

「免疫応答」は、炎症、外傷、免疫異常症、あるいは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。

40

【0088】

「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IGSF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IGSF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。

【0089】

用語「マイクロアレイ」は、基板上の複数のポリヌクレオチド、ポリペプチドまたはその

50

他の化合物の構成を指す。

【0090】

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

【0091】

「調節(する)」の語は、IGSFPの活性の変化を指す。例えば、調節によって、IGSFPのタンパク質活性、或いは結合特性、またはその他の生物学的特性、機能的特性或いは免疫学的特性の変化が起きうる。

【0092】

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

10

【0093】

「機能的に連結した」は、第1の核酸配列と第2の核酸配列が機能的な関係にある状態を指す。例えば、プロモーターがコード配列の転写または発現に影響を及ぼす場合には、そのプロモーターはそのコード配列に機能的に連結している。機能的に連結したDNA配列は非常に近接するか、或いは連続的に隣接し得る。また、2つのタンパク質コード領域を結合するために必要な場合は、同一のリーディングフレーム内に在り得る。

20

【0094】

「ペプチド核酸(PNA)」は、末端がリジンで終わ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を有する、アンチセンス分子または抗遺伝子剤を指す。末端のリジンは、組成物に溶解性を与える。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伸長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。

【0095】

IGSFPの「翻訳後修飾」には、脂質化、グリコシル化、リン酸化、アセチル化、ラセミ化、蛋白分解性切断及びその他の当分野で既知の修飾を含まれ得る。これらのプロセスは、合成或いは生化学的に生じ得る。生化学的修飾は、IGSFPの酵素環境に依存し、細胞の種類によって異なることとなる。

30

【0096】

「プローブ」とは、同一配列、対立遺伝子核酸配列、或いは関連する核酸配列の検出に用いる、IGSFP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」とは、相補的な塩基対を形成して標的のポリヌクレオチドにアニーリング可能な、通常はDNAオリゴヌクレオチドである短い核酸である。プライマーは次に、DNAポリメラーゼ酵素によって標的DNA鎖に沿って延長し得る。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。

40

【0097】

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または少なくとも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。

【0098】

50

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.他 (1989) Molecular Cloning: A Laboratory Manual, 第2版, 1-3巻、Cold Spring Harbor Press, Plainview NY、Ausubel, F.M.他, (1987) Current Protocols in Molecular Biology, Greene Publi. Assoc. & Wiley-Intersciences, New York NY、Innis他 (1990) PCR Protocols, A Guide to Methods and Applications Academic Press, San Diego CA等を参照されたい。PCRプライマー対は、その目的のためのコンピュータプログラム、例えばPrimer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA) を用いるなどして既知の配列から得ることができる。

【0099】

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既知の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、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オリゴヌクレオチド及び最大5,000ヌクレオチドまでの大きめのポリヌクレオチドとオリゴヌクレオチド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 (Whitehead Institute/MIT Center for Genome Research (マサチューセッツ州ケンブリッジ) より入手可能) によって、ユーザーは、プライマー結合部位として避けたい配列を指定できる「ミスプライミングライブラリ」を入力できる。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、それぞれの情報源から得てユーザー固有のニーズを満たすように修正し得る)。PrimerGenプログラム (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 (英国ケンブリッジ) 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、固有なもの、及び保存されたもの双方のオリゴヌクレオチドとポリヌクレオチド断片と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。

【0100】

「組換え核酸」は天然の配列ではなく、2つ以上の配列の離れたセグメントを人工的に組み合わせた配列である。この人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群の人為的操作によって、例えばSambrookの文献 (前出) に記載されているような遺伝子工学的手法によって達成する。組換え核酸の語は、単に核酸の一部が付加、置換または欠失により改変された核酸も含む。しばしば組換え核酸には、プロモーター配列に機能的に連結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。

【0101】

或いはこのような組換え核酸は、ウイルスベクターの一部であって、例えばワクシニアウイルスに基づくものであり得る。そのようなワクシニアウイルスは哺乳動物に接種され、その組換え核酸が発現されて、その哺乳動物内で防御免疫応答を誘導するように使用することができる。

【0102】

「調節エレメント」は、通常は遺伝子の非翻訳領域に由来する核酸配列であり、エンハンサー、プロモーター、イントロン及び5'及び3'の非翻訳領域 (UTR) を含む。調節エレ

10

20

30

40

50

メントは、転写、翻訳またはRNA安定性を調節する宿主タンパク質またはウイルスタンパク質と相互作用する。

【0103】

「レポーター分子」は、核酸、アミノ酸または抗体の標識化に用いられる化学的または生化学的な部分である。レポーター分子としては、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を含む。

DNA配列に対する「RNA等価物」は、基準となるDNA配列と同じ直鎖の核酸配列から構成されるが、全ての窒素性塩基のチミンがウラシルで置換され、糖鎖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なる。

10

【0104】

用語「サンプル」は、その最も広い意味で用いられている。IGSFP、IGSFPをコードする核酸、またはその断片を含むと推定されるサンプルは、体液と、細胞からの抽出物や細胞から単離された染色体や細胞内小器官、膜と、細胞と、溶液中に存在するまたは基板に固定されたゲノムDNA、RNA、cDNAと、組織と、組織プリント等を含み得る。

【0105】

用語「特異的結合」及び「特異的に結合する」は、タンパク質若しくはペプチドと、アゴニスト、抗体、アンタゴニスト、小分子、若しくは任意の天然若しくは合成の結合組成物との間の相互作用を指す。この相互作用は、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構造が存在するか否かに依存する。例えば、抗体がエピトープ「A」に対して特異的である場合、結合していない標識した「A」及び抗体を含む反応液に、エピトープAを含むポリペプチド或いは結合していない無標識の「A」が存在すると、抗体と結合する標識Aの量が減少する。

20

【0106】

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

【0107】

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

30

【0108】

「基板」は、任意の好適な固体或いは半固体の支持物を指し、膜及びフィルター、チップ、スライド、ウエハ、ファイバー、磁気または非磁気ビーズ、ゲル、チューブ、プレート、ポリマー、微小粒子、毛細管が含まれる。基板は、壁、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、基板にはポリヌクレオチドやポリペプチドが結合する。

【0109】

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

40

【0110】

「形質転換(transformation)」とは、外来DNAが受容細胞に導入されるプロセスのことである。形質転換は、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宿主細胞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或いはウイルス感染、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。用語「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、限られた時間に一過的に導入DNA若しくは

50

導入RNAを発現する細胞も含まれる。

【0111】

ここで用いる「遺伝形質転換体(transgenic organism)」とは任意の生物体であり、限定するものではないが動植物を含み、生物体の1個以上の細胞が、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トランスジェニック(transgenic)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での感染によって行う。あるいは、核酸をレンチウイルスベクターのような組換えウイルスベクターで感染させて導入することができる(Lois, C.他.(2002)Science 295:868-872)。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をこのような生物体に移入する技術はよく知られており、前出のSambrook 他 (1989)等の参考文献に記載されている。

10

【0112】

特定の核酸配列の「変異体/変異配列」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%の配列同一性を有する核酸配列であると定義する。定義づけには、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%またはそれ以上の配列同一性を示し得る。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の有意な同一性を有し得るが、mRNAプロセッシング中のエキソンの選択的スプライシングによって通常、より多くまたはより少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、種によって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、相互に有意なアミノ酸同一性を有する。多型性変異体は、所与の種の個体間での特定遺伝子のポリヌクレオチド配列中での変異である。多型変異配列はまた、ポリヌクレオチド配列の1つのヌクレオチド塩基が異なる「1塩基多型性」(SNP)も含み得る。SNPの存在は、例えば特定の個体群、病状または病状性向を示し得る。

20

30

【0113】

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%の配列同一性を有するポリペプチド配列として定義される。定義づけには、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

【0114】

(発明)

本発明は、新規のヒト免疫グロブリンスーパーファミリータンパク質(IGSFP)及びIGSFPをコードするポリヌクレオチドの発見に基づき、これらの組成物を利用した免疫系の疾患、神経の疾患、発生または発達障害、筋疾患および細胞増殖異常の診断、治療、及び予防に関する。

【0115】

表1は、本発明の完全長ポリヌクレオチド配列及びポリペプチド配列の命名の概略である。各ポリヌクレオチド及びその対応するポリペプチドは、1つのIncyteプロジェクト識別

50

番号 (IncyteプロジェクトID) と相関する。各ポリペプチド配列は、ポリペプチド配列識別番号 (ポリペプチドSEQ ID NO) とIncyteポリペプチド配列番号 (IncyteポリペプチドID) によって表示した。各ポリヌクレオチド配列は、ポリヌクレオチド配列識別番号 (ポリヌクレオチドSEQ ID NO) とIncyteポリヌクレオチド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。列6は本発明のポリペプチドおよびポリヌクレオチド配列に相当する物理的な完全長クローンのIncyte ID 番号を示す。完全長のクローンは列3に示すポリペプチド配列に少なくとも95%の配列同一性を有するポリペプチドをコードする。

【0116】

表2は、GenBankタンパク質 (genpept) データベースとPROTEOMEデータベースとに対するBLAST分析で同定した、本発明のポリペプチド群に相同な配列群を示す。列1および列2はそれぞれ、本発明の各ポリペプチドに対するポリペプチド配列識別番号 (ポリペプチドSEQ ID NO:) とそれに対応するIncyteポリペプチド配列番号 (IncyteポリペプチドID) を示す。列3は、GenBankの最も近い相同体のGenBank識別番号 (Genbank ID NO:) と最も近いPROTEOME データベース相同体のPROTEOME データベース識別番号 (PROTEOME ID NO:) を示す。列4は、各ポリペプチドとその相同体との間の一致について確率スコアを示す。列5は、GenBankとPROTEOME データベースの相同体の注釈を示し、更に該当箇所には関連する引用文献も示す。これらを引用することを以って本明細書の一部とする。

【0117】

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ポリペプチドID) を示す。列3は、各ポリペプチドのアミノ酸残基数を示す。列4および列5はそれぞれ、GCG配列分析ソフトウェアパッケージのMOTIFSプログラム (Genetics Computer Group, Madison WI) によって決定された、リン酸化およびグリコシル化の可能性のある部位を示す。列6は、シグネチャ配列、ドメイン、およびモチーフを有する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。

【0118】

表2および3は共に、本発明の各々のポリペプチドの特性を要約しており、それら特性が、請求の範囲に記載されたポリペプチドが免疫グロブリンスーパーファミリータンパク質であることを確立している。例えば、SEQ ID NO:2はマウスFca/m受容体 (GenBank ID g11071950) とQ34残基からP563残基まで50%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 $9.6e-121$ であるが、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す。SEQ ID NO:2もまた、免疫グロブリンドメインを有するが、これは、隠れマルコフモデル (HMM) を基にした保存されたタンパク質ファミリードメインのPFAMデータベースにおいて、統計的に有意な一致を検索して決定された。(表3参照)。付加的なBLAST解析よりのデータは、SEQ ID NO:2が免疫グロブリンである、さらに実証的な証拠を提供する。別の例において、SEQ ID NO:3はL30残基からV176残基まで表面タンパク質MCA-32 (GenBank ID g1136501) に40%の同一性を有するが、これはBasic Local Alignment Search Tool (BLAST) によって同定される。(表2参照)。BLAST確率スコアは $6.9e-35$ 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:3もまた、免疫グロブリンドメインを有するが、これは、隠れマルコフモデル (HMM) を基にした保存されたタンパク質ファミリードメインのPFAMデータベースにおいて、統計的に有意な一致を検索して決定された。(表3参照)。BLIMPS、MOTIFS及び付加的なBLAST解析よりのデータは、SEQ ID NO:3が表面タンパク質である、さらに実証的な証拠を提供する。また他の例として、SEQ ID NO:8は細胞表面分子Ly-9 (GenBank ID g10197717) とM1残基からS433残基まで86%同一であることが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 $7.4e-191$ であるが、これは観測され

10

20

30

40

50

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す。SEQ ID NO:8はまた、免疫グロブリンドメインを有することが、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)。付加的BLAST解析よりのデータは、SEQ ID NO:8が免疫グロブリンスーパーファミリーのメンバーである細胞表面分子であるさらに実証的な証拠を提供する。別の例において、SEQ ID NO:11はN43残基からQ604残基までヒトのNEPH1(GenBank ID g14572521)に52%の同一性を有するが、これはBasic Local Alignment Search Tool (BLAST)によって同定される。(表2参照)。BLAST確率スコアは $5.4e-158$ であるが、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す。PROTEOMEデータベースを使ったBLAST解析によって示されたように、SEQ ID NO:11は原形質膜に局在し、免疫グロブリンドメインを含むヒトタンパク質に相同する。そしてグリコシルホスファチジルイノシトール(GPI)アンカー神経細胞接着分子(PROTEOME ID 598720|FLJ10845)であるオピオイド結合細胞接着分子の領域に対して類似性の低い領域を有する。SEQ ID NO:11はまた、腎臓濾過障壁の機能または発達である役割を果たしている可能性のある腎臓の糸球体で発現する免疫グロブリンスーパーファミリーのメンバーであるヒトネフリンに対して相通的である。ネフリン遺伝子の突然変異は先天性ネフローゼ症候群を引き起こす(PROTEOME ID 340970|NPHS1)。SEQ ID NO:11もまた、免疫グロブリンドメインを有するが、これは、隠れマルコフモデル(HMM)を基にした保存されたタンパク質ファミリドメインの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、および付加的BLAST解析から得られたデータによって、SEQ ID NO:11が免疫グロブリンスーパーファミリーのメンバーであるさらに実証的な証拠を提供する。SEQ ID NO:1、SEQ ID NO:4-7、SEQ ID NO:9-10、およびSEQ ID NO:12については、同様の方法で分析し、注釈を付けた。SEQ ID NO:1-12の解析用のアルゴリズム及びパラメータを表7に記載した。

10

20

30

40

50

【0119】

表4に示すように、本発明の完全長ポリヌクレオチド配列は、cDNA配列またはゲノムDNA由来のコード(エキソン)配列を用いて、或いはこれら2種類の配列を任意に組み合わせて構築(assemble)した。列1は本発明の各ポリヌクレオチドに対する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)、対応するIncyteポリヌクレオチドコンセンサス配列番号(Incyte ID)、および塩基対での各ポリヌクレオチド配列の長さを示している。列2は、本発明の完全長ポリヌクレオチド配列の構築に使われたcDNA配列および/またはゲノム配列のヌクレオチド開始位置(5')と停止位置(3')、およびSEQ ID NO:13-24を同定するか、またはSEQ ID NO:13-24と、関連するポリヌクレオチド配列とを区別する技術(例えば、ハイブリダイゼーション技術または増幅技術)に有用なポリヌクレオチド配列の断片のヌクレオチド開始位置(5')と終了位置(3')を示す。

【0120】

表4の列2で記述されたポリヌクレオチド断片は特に、例えば組織特異的cDNAライブラリあるいはプールしたcDNAライブラリに由来するIncyte cDNAを指す場合もある。或いは列2のポリヌクレオチド断片は、完全長ポリヌクレオチド配列の構築に寄与したGenBank cDNAまたはESTを指す場合もある。さらに、列2のポリヌクレオチド断片は、ENSEMBL(The Sanger Centre、(英国ケンブリッジ))データベースから由来した配列を同定し得る(即ち「ENST」命名を含む配列)。或いは、列2で記述されたポリヌクレオチド断片は、NCBI RefSeq Nucleotide Sequence Records データベースから由来する場合もあり(即ち「NM」または「NT」の命名を含む配列)、またNCBI RefSeq Protein Sequence Recordsから由来する場合もある(即ち「NP」の命名を含む配列)。または列2のポリヌクレオチド断片は、「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意味する場合がある。例えば、FL_XXXXXX_N₁_N₂_YYYYY_N₃_N₄と同定されるポリヌクレオチドは、アルゴリズムが適用される配列のクラスターの識別番号がXXXXXXであり、アルゴリズムにより生成される予測の番号がYYYYYであり、(もし存在すれば)N_{1,2,3...}が解析中に手動で編集された可能性のある特定のエキ

ソンであるような「スティッチされた(stitched)」配列である(実施例5参照)。または、列2のポリヌクレオチド断片は「エキソストレッチ(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場合もある。例えば、FLXXXXXX_gAAAAA_gBBBBB_1_Nとして同定されるポリヌクレオチド配列は、「ストレッチされた」配列である。XXXXXXはIncyteプロジェクト識別番号、gAAAAAは「エキソストレッチ」アルゴリズムを適用したヒトゲノム配列のGenBank識別番号、gBBBBBは一番近いGenBankタンパク質相同体のGenBank識別番号または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)。あるRefSeq配列が「エキソストレッチ」アルゴリズムのためのタンパク質相同体として使用された場合には、RefSeq識別子(「NM」、「NP」、または「NT」によって表される)が、GenBank識別子(即ち、gBBBBB)の代わりに使用される場合もある。

10

【0121】

或いは、接頭コードは、手動で編集された構成配列、ゲノムDNA配列から予測された構成配列、または組み合わされた配列解析方法から由来する構成配列を同定する。次の表は、構成配列の接頭コードと、接頭コードに対応する配列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	解析タイプやプログラムの例
GNN GFG ENST	例えば、GENSCAN (Stanford University, CA, USA) または FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK)を用いた、ゲノム配列群からのエキソン予測
GBI	手動で編集された、ゲノム配列群の解析
FL	スティッチまたはストレッチゲノム配列(実施例5参照)
INCY	EST 配列群の、ゲノムへのマッピングからの、完全長転写物とエキソンとの予想。エキソンと転写物を予想するために、ゲノム位置と EST 組成のデータが組み合わされる。

20

30

【0122】

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に表4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、それに関連するIncyte cDNA識別番号は示さなかった。

【0123】

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

40

【0124】

本発明はまた、IGSFPの変異体も含む。好適なIGSFPの変異体は、IGSFPアミノ酸配列に対して、少なくとも約80%、あるいは少なくとも約90%、さらには少なくとも約95%のアミノ酸配列同一性を有し、IGSFPの機能的または構造的特徴を少なくとも1つ含む変異体である。

50

【0125】

本発明はまた、IGSFPをコードするポリヌクレオチドを提供する。特定の実施例において、本発明は、IGSFPをコードするSEQ ID NO:13-24からなる一群から選択された配列を含むポリヌクレオチド配列を提供する。SEQ ID NO:13-24のポリヌクレオチド配列には、配列表に示したように等価RNA配列をも含むが、そこでは窒素塩基チミンの出現はウラシルで置換され、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で構成される。

【0126】

本発明はまた、IGSFPをコードするポリヌクレオチド配列の変異配列を含む。詳細には、このようなポリヌクレオチド配列の変異配列は、IGSFPをコードするポリヌクレオチド配列と少なくとも約70%のポリヌクレオチド配列同一性、或いは少なくとも約85%のポリヌクレオチド配列同一性、更には少なくとも約95%ものポリヌクレオチド配列同一性を有する。本発明の特定の実施形態は、SEQ ID NO:13-24からなる一群から選択された核酸配列と少なくとも約70%のポリヌクレオチド配列同一性、或いは少なくとも約85%のポリヌクレオチド配列同一性、更には少なくとも約95%ものポリヌクレオチド配列同一性を有するSEQ ID NO:13-24からなる一群から選択された配列を含むポリヌクレオチド配列の変異配列を提供する。上記したポリヌクレオチド変異配列は何れも、IGSFP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。

【0127】

さらに、或いは別法では、本発明のポリヌクレオチド変異体は、IGSFPをコードするポリヌクレオチド配列のスプライス変異体である。スプライス変異体はIGSFPをコードするポリヌクレオチド配列との有意の配列同一性を持つ部分複数を有し得るが、mRNAプロセッシング中のエキソン群の選択的スプライシングによる、配列ブロック群の付加または欠失により、通常、より多くまたはより少数のポリヌクレオチドを有することになる。スプライス変異体はその全長についてIGSFPをコードするポリヌクレオチド配列とのポリヌクレオチド配列同一性が約70%未満、あるいは約60%、あるいは約50%未満でありえるが、そのスプライス変異体のある部分は、IGSFPをコードするポリヌクレオチド配列の各部分とのポリヌクレオチド配列同一性が少なくとも約70%、あるいは少なくとも約85%、あるいは少なくとも約95%、あるいは100%になる。例えば、SEQ ID NO:14の配列を含むポリヌクレオチドはSEQ ID NO:24の配列を含むポリヌクレオチドのスプライス変異体であり、SEQ ID NO:16の配列を含むポリヌクレオチドはSEQ ID NO:17の配列を含むポリヌクレオチドのスプライス変異体である。上記したスプライス変異配列は何れも、IGSFPの機能的或いは構造的特徴の少なくとも1つを有する或るアミノ酸配列をコードし得る。

【0128】

遺伝暗号の縮重により、IGSFPをコードする種々のポリヌクレオチド配列が作り出され、中には、既知のいかなる天然遺伝子のポリヌクレオチド配列群とも最小の類似性しか有しない配列もあることは、当業者には理解されよう。したがって本発明は、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のパリエーションを網羅し得る。これらの組み合わせは、天然IGSFPのポリヌクレオチド配列に適用される標準的なトリプレット遺伝暗号を基に作られる。また、このような全ての変異が明確に開示されていると考慮されたい。

【0129】

IGSFPをコードするヌクレオチド配列及びその変異配列は一般に、好適に選択されたストリンジェントな条件下で、天然のIGSFPのヌクレオチド配列とハイブリダイズ可能であるが、非天然のコドンを含めるなどの実質的に異なった使い方のコドンを含むIGSFP或いはその誘導体をコードするヌクレオチド配列を作るとは有利となり得る。宿主が特定のコドンを利用する頻度に基づいて、特定の真核宿主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。コードされるアミノ酸配列を改変せずに、IGSFP及びその誘導体をコードするヌクレオチド配列を実質的に改変する別の理由には、天然配列から作られる転写物より例えば長い半減期など好ましい特性を備えるRNA転写物を作ることもある。

10

20

30

40

50

【0130】

本発明はまた、IGSFP及びその誘導体をコードするDNA配列またはそれらの断片を完全に合成化学によって作り出すことも含む。作製後にこの合成配列を、当分野で良く知られた試薬を用いて、種々の入手可能な発現ベクター及び細胞系の何れの中にも挿入可能である。更に、合成化学を用いて、IGSFPまたはその任意の断片をコードする配列の中に突然変異を導入することも可能である。

【0131】

更に本発明には、種々のストリンジェンシー条件下で、請求項に記載されたポリヌクレオチド配列に、特に、SEQ ID NO:13-24及びそれらの断片群にハイブリダイズ可能なポリヌクレオチド配列群が含まれる（例えば、Wahl, G.M.および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。

【0132】

DNAシーケンシングの方法は当分野では公知であり、本発明のいずれの実施例も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1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ラーゼを併用することができる。好適には、MICROLAB2200液体転移システム (Hamilton, Reno, NV)、PTC200サーマルサイクラー (MJ Research, Watertown MA) 及びABI CATALYST 800サーマルサイクラー (Applied Biosystems) 等の装置を用いて配列の準備を自動化する。次に、ABI 373或いは377 DNAシーケンシングシステム (Applied Biosystems)、MEGABACE 1000 DNAシーケンシングシステム (Molecular Dynamics, Sunnyvale CA) または当分野でよく知られている他の方法を用いてシーケンシングを行う。結果として得られた配列を当分野でよく知られている種々のアルゴリズムを用いて分析する (Ausubel, F.M. (1997) *Short Protocols in Molecular Biology*, John Wiley & Sons, New York NY, unit 7.7、Meyers, R.A. (1995) *Molecular Biology and Biotechnology*, Wiley VCH, New York NY, 856-853ページ等を参照)。

【0133】

IGSFP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エレメント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 (例えば、Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322を参照)。別の方法に逆PCR法があり、これは広範な方向に伸長させたプライマーを用いて環状化した鋳型から未知の配列を増幅する方法である。鋳型は、既知のゲノム遺伝子座及びその周辺の配列を含む制限酵素断片から得る (例えばTriglia, T. 他 (1988) *Nucleic Acids Res.* 16:8186を参照)。第3の方法としてキャプチャPCR法があり、これはヒト及び酵母菌人工染色体DNAの既知の配列に隣接するDNA断片をPCR増幅する方法に参与している (例えばLagerstrom, M. 他 (1991) *PCR Methods Applic.* 1:111119を参照)。この方法では、PCRを行う前に複数の制限酵素の消化及びライゲーション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。未知の配列群を検索するために用い得る他の複数の方法も当分野で既知である (例えばParker, J.D. 他 (1991) *Nucleic Acids Res.* 19:30553060を参照)。更に、PCR、ネステッドプライマー及びPromoterFinder (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。

【0134】

完全長cDNA群をスクリーニングする際は、より大きなcDNA群を含むようにサイズ選択されたライブラリ群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、遺伝子群の5'領域を有する配列をしばしば含んでおり、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。

【0135】

市販のキャピラリー電気泳動システムを用いて、シークエンシング産物またはPCR産物のサイズを分析し、またはそのヌクレオチド配列を確認することができる。具体的には、キャピラリーシークエンシングは、電気泳動による分離のための流動性ポリマーと、4つの異なるヌクレオチドに特異的であるような、レーザーで刺激される蛍光色素と、発光された波長の検出に利用するCCDカメラとを有し得る。出力/光の強度は、適切なソフトウェア (Applied Biosystems社のGENOTYPER、SEQUENCE NAVIGATOR等) を用いて電気信号に変換し得る。サンプルのロードからコンピュータ分析及び電子データ表示までの全プロセスがコンピュータ制御可能である。キャピラリー電気泳動法は、特定のサンプルに少量しか存在しないようなDNA小断片のシークエンシングに特に適している。

【0136】

本発明の別の実施態様では、適切な宿主細胞内で、IGSFP、その断片または機能的等価物を発現させる組換えDNA分子群内で、IGSFPをコードするポリヌクレオチド配列群またはその断片群をクローニングし得る。遺伝暗号固有の縮重により、実質的に同じ或いは機能的に等価のアミノ酸配列をコードする別のDNA配列が作られ得、これらの配列をIGSFPの発現に利用可能である。

【0137】

種々の目的で、IGSFPをコードする配列を改変するために、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再構築による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を介した部位特異的変異誘導を利用して、新規な制限部位の作製、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起こす突然変異を導入し得る。

【0138】

本発明のヌクレオチドは、MOLECULARBREEDING (Maxygen Inc., Santa Clara CA、米国特許第5,837,458号、Chang, C.-C.他 (1999) Nat. Biotechnol. 17:793-797、Christians, F.C.他 (1999) Nat. Biotechnol. 17:259-264、Cramer, A. 他 (1996) Nat. Biotechnol. 14:315-319に記載) 等のDNAシャッフリング技術の対象となり、IGSFPの生物学的特性、例えば生物活性、酵素活性、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を介する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、所望の特性を持つ遺伝子変異配列群を同定する、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の、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。かくして、「人工的な」育種及び急速な分子進化によって遺伝的多様性が作られる。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、所与遺伝子の断片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の断片と組み換え、それによって天然に存在する複数の遺伝子の遺伝的多様性を、定方向の、制御可能な方法で最大化させることができる。

10

20

30

40

50

【0139】

別の実施例によれば、IGSFPをコードする配列は、当分野で周知の化学的方法を用いて、全体或いは一部が合成可能である（例えば、Caruthers, M.H.他（1980）Nucl. Acids Res. Symp. Ser 7:215-223、Horn, T.他（1980）Nucl. Acids Res. Symp. Ser7.225-232を参照）。別法として、化学的方法を用いてIGSFP自体またはその断片を合成することが可能である。例えば、種々の液相または固相技術を用いてペプチド合成を行うことができる（Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, 55-60ページ、Roberge, J.Y. 他（1995）Science 269:202-204等を参照）。自動合成はABI 431Aペプチドシンセサイザ（Applied Biosystems）を用いて達成し得る。更にIGSFPのアミノ酸配列または任意のその一部を用い、直接合成の際の改変により、及び/または化学的方法を用いた他のタンパク質からの配列または任意のその一部との組み合わせにより、天然ポリペプチドの配列を有するポリペプチドまたは変異体ポリペプチドを作製することが可能である。

10

【0140】

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

【0141】

生物学的に活性なIGSFPを発現させるために、IGSFPをコードするヌクレオチド配列またはその誘導体を好適な発現ベクターに挿入する。この発現ベクターは、好適な宿主に挿入されたコーディング配列の転写及び翻訳の調節に必要なエレメントを含む。これらのエレメントには、ベクター及びIGSFPをコードするポリヌクレオチド配列における、エンハンサー、構成型及び発現誘導型のプロモーター、5'及び3'の非翻訳領域などの調節配列が含まれる。このようなエレメントは、強度及び特異性が様々である。特定の開始シグナル類を用いて、IGSFPをコードする配列群の、より効果的な翻訳を達成できる。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。IGSFPをコードする配列群、その開始コドン、および上流の調節配列群が、好適な発現ベクターに挿入された場合は、更なる転写制御シグナルや翻訳制御シグナルは必要ないこともある。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因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。外来性の翻訳エレメント及び開始コドンは、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. 20:12 5162を参照）。

20

30

【0142】

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

40

【0143】

種々の発現ベクター/宿主系を利用して、IGSFPをコードする配列の保持及び発現が可能である。限定するものではないがこのような発現ベクター/宿主系には、組換えバクテリオファージ、プラスミドまたはコスミドDNA発現ベクターで形質転換させた細菌などの微生物等や、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プラスミドまたはpBR322プラスミド）で形質転換させた植物細胞系または動物

50

細胞系がある（前出のSambrook、前出のAusubel、Van Heeke, G.及びS.M. Schuster (1989) J. Biol. Chem. 264:5503-5509、Engelhard, E.K. 他 (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227、Sandig, V. 他 (1996) Hum. Gene Ther. 7:1937-1945、Takamatsu, N. (1987) EMBO J. 6:307-311、; 『マグローヒル科学技術年鑑』(The McGraw Hill Yearbook of Science and Technology) (1992) McGraw Hill New York NY, 191-196ページ、Logan, J.及びT. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659、Harrington, J.J. 他 (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送達することができる (Di Nicola, M. 他 (1998) Cancer Gen. Ther. 5(6):350-356、Yu, M. 他 (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344、Buller, R.M. 他 (1985) Nature 317(6040):813-815、McGregor, D.P. 他 (1994) Mol. Immunol. 31(3):219-226、Verma, I.M.及びN. Somia (1997) Nature 389:239-242等を参照)。本発明は使用される宿主細胞によって限定されるものではない。

10

【0144】

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、IGSFP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、IGSFPをコードするポリヌクレオチド配列群の慣例的なクローニング、サブクローニング、増殖には、PBLUESCRIPT(Stratagene, La Jolla CA)またはpSPORT1 プラスミド(Life Technologies)などの、多機能の大腸菌ベクターを用い得る。ベクターのマルチクローニング部位にIGSFPをコードする配列を連結反応するとlacZ 遺伝子が破壊され、組換え分子を含む形質転換された細菌の同定のための比色スクリーニング法が可能となる。更にこれらのベクターは、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レスキュー、入れ子状態の欠失の生成にも有用であろう (Van Heeke, G.及びS.M. Schuster (1989) J. Biol. Chem. 264:5503-5509等を参照)。例えば、抗体の産生のためなどに多量のIGSFPが必要な場合は、IGSFPの高レベル発現を誘導するベクターが使用できる。例えば、強力な誘導SP6バクテリオファージプロモーターまたは誘導T7バクテリオファージプロモーターを有するベクターが使用できる。

20

【0145】

IGSFPの産生には、酵母発現系の使用が可能である。因子、アルコールオキシダーゼ、PGHプロモーター等の構成型或いは誘導型のプロモーターを含む多数のベクターが、出芽酵母菌 (*Saccharomyces cerevisiae*) またはピキア酵母 (*Pichia pastoris*) に使用可能である。更に、このようなベクターは、発現したタンパク質の分泌か細胞内への保持のどちらかを指示する。また、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ことを可能にする (例えば前出のAusubel, 1995、Bitter, G.A. 他 (1987) Methods Enzymol. 153:516544、Scorer, C.A. 他 (1994) Bio/Technology 12:181-184を参照)。

30

【0146】

植物系を使用してIGSFPを発現することも可能である。IGSFPをコードする配列の転写は、ウイルスプロモーター、例えば単独或いはTMV (Takamatsu, N. (1987) EMBO J 6:307-311) 由来のオメガリーダー配列と組み合わせて用いられるようなCaMV由来の35S及び19Sプロモーターによって促進される。或いは、RUBISCOの小サブユニット等の植物プロモーターまたは熱ショックプロモーターを用いてもよい (Coruzzi G 他 (1984) EMBO J. 3:1671-1680、Broglie R 他 (1984) Science 224:838-843、Winter, J 他 (1991) Results Probl. Cell Differ. 17:85-105等を参照)。これらの作製物は、直接DNA形質転換にまたは病原体を媒介とする形質移入によって、植物細胞内に導入可能である (『マグローヒル科学技術年鑑』(The McGraw Hill Yearbook of Science and Technology) (1992) McGraw Hill New York NY, 191-196ページ等を参照)。

40

【0147】

哺乳動物細胞においては、多数のウイルスベースの発現系を利用し得る。アデノウイルスが発現ベクターとして用いられる場合、後発プロモーター及び3連リーダー配列からなる

50

アデノウイルス転写物/翻訳複合体にIGSFPをコードする配列を連結し得る。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IGSFPを発現する感染ウイルスを得ることが可能である。(例えば、Logan, J.及びT. Shenk (1984) Proc. Natl. Acad. Sci. U.S.A. 81:3655-3659等を参照)。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。

【0148】

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の断片を送達することもできる。治療のために約6kb~10MbのHACを作製し、従来の送達方法(リポソーム、ポリカチオンアミノポリマー、またはベシクル)で送達する(例えばHarrington, J.J.他(1997) Nat. Genet. 15:345-355を参照)。

【0149】

哺乳動物系の組換えタンパク質の長期にわたる産生のためには、株化細胞におけるIGSFPの安定した発現が望ましい。例えば、IGSFPをコードする配列を細胞株に形質転換するために、発現ベクター類と、同じベクター上の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得る。用いる発現ベクターは、ウイルス起源の複製要素、および/または内因性の発現要素を持ち得る。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。

【0150】

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

【0151】

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、IGSFPをコードする配列がマーカー遺伝子配列の中に挿入された場合、IGSFPをコードする配列を含む形質転換された細胞は、マーカー遺伝子機能の欠落により同定可能である。または、単一プロモーターの制御下で、或るマーカー遺伝子を、IGSFPをコードする1配列とタンデムに配置することもできる。誘導または選択に応答したマーカー遺伝子の発現は通常、タンデム遺伝子も発現していることを意味する。

【0152】

一般に、IGSFPをコードする核酸配列を持ち、IGSFPを発現する宿主細胞は、当業者に周知

の種々の方法を用いて特定できる。これらの方法には、DNA - DNA或いはDNA - RNAハイブリダイゼーションや、PCR法、核酸或いはタンパク質配列の検出及び/または数量化のための膜系、溶液ベース、或いはチップベースの技術を含むタンパク質生物学的試験法または免疫学的アッセイが含まれるが、これらに限定されるものではない。

【0153】

特異的なポリクローナル抗体またはモノクローナル抗体のどちらかを用いる、IGSFPの発現の検出及び計測のための免疫学的な方法は、当分野で既知である。このような技法には、酵素結合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) などがある。IGSFP上の2つの非干渉エピトープに反応するモノクローナル抗体を用いた、2部位のモノクローナルベース イムノアッセイ (two-site, monoclonal-based immunoassay) が好ましいが、競合の結合アッセイも用いることもできる。これらのアッセイ及びこれ以外のアッセイは、当分野で公知である (Hampton, R. 他 (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN, Sect. IV、Coligan, J. E. 他 (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY、Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ等を参照)。

10

【0154】

多岐にわたる標識方法及び抱合方法が、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。IGSFPをコードするポリヌクレオチドに関連する配列を検出するための、標識されたハイブリダイゼーションプローブ或いはPCRプローブを生成する方法には、オリゴ標識化、ニックトランスレーション、エンドラベリング (末端標識化)、または標識されたヌクレオチドを用いるPCR増幅が含まれる。別法として、IGSFPをコードする配列、またはその任意の断片をmRNAプローブの生成の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)、および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、磁気粒子の他、放射性核種、酵素、蛍光剤、化学発光剤、発色剤等がある。

20

30

【0155】

IGSFPをコードするヌクレオチド配列で形質転換された宿主細胞は、細胞培地でのこのタンパク質の発現及び回収に好適な条件下で培養され得る。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。IGSFPをコードするポリヌクレオチドを含む発現ベクターは、原核細胞膜及び真核細胞膜を透過するIGSFPの分泌を誘導するシグナル配列を含むように設計できることは、当業者には理解されよう。

【0156】

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」形を切断する翻訳後のプロセッシングを利用して、タンパク質の標的への誘導、折りたたみ及び/または活性を特定することが可能である。翻訳後の活性のための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。

40

【0157】

本発明の別の実施例では、IGSFPをコードする天然の核酸配列、変更された核酸配列、または組換えの核酸配列を上記した任意の宿主系の融合タンパク質の翻訳となる異種配列に

50

連結させ得る。例えば、市販の抗体によって認識できる異種部分を含むキメラIGSFPタンパク質が、IGSFP活性のインヒビターに対するペプチドライブラリのスクリーニングを促進し得る。また、異種タンパク質部分及び異種ペプチド部分も、市販されている親和性基質を用いて融合タンパク質の精製を促進し得る。限定されるものではないがこのような部分には、グルタチオンSトランスフェラーゼ(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されているモノクローナル抗体及びポリクローナル抗体を用いて、融合タンパク質の免疫親和性精製を可能にする。また、IGSFPをコードする配列と異種タンパク質配列との間にあるタンパク質分解切断部位を融合タンパク質が含むように遺伝子操作すると、IGSFPが精製の後に異種部分から切断され得る。融合タンパク質の発現及び精製方法は、前出の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。

10

【0158】

本発明の別の実施例では、TNTウサギ網状赤血球可溶化液またはコムギ胚芽抽出系(Promega)を用いて *in vitro* で放射能標識したIGSFPの合成が可能である。これらの系は、T7、T3またはSP6プロモーターと機能的に連結したタンパク質コード配列の転写及び翻訳をカッ

20

【0159】

本発明のIGSFPまたはその断片を用いて、IGSFP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、IGSFP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。

【0160】

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IGSFP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、この化合物はIGSFPが結合する天然受容体、或いは、少なくともこの受容体の或る断片、例えばリガンド結合部位など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IGSFP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、あるいは大腸菌からの細胞が含まれる。次に、IGSFPを発現する細胞、またはIGSFPを含む細胞膜分画を試験化合物と接触させて、IGSFPまたは試験化合物の何れかの、結合または活性の刺激あるいは阻害を分析する。

30

【0161】

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体に固定されたIGSFPと混合するステップと、IGSFPとこの化合物との結合を検出するステップを含み得る。別法では、アッセイ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

【0162】

本発明のIGSFPまたはその断片を用いて、IGSFPの活性を調整する化合物をスクリーニング

50

することが可能である。このような化合物としては、アゴニスト、アンタゴニスト、部分的アゴニストまたは逆アゴニストなどが含まれ得る。一実施例では、IGSFPの活性が許容される条件下でアッセイを実施し、そのアッセイでは少なくとも1つの試験化合物をIGSFPと混合し、試験化合物の存在下でIGSFPの活性を試験化合物不在下でのIGSFPの活性と比較する。試験化合物の存在下でのIGSFPの活性の変化は、IGSFPの活性を調整する化合物の存在を示唆する。別の実施態様において、試験化合物をIGSFPの活性に適した条件下でIGSFPを含む *in vitro* または無細胞再構成系と結合させてアッセイを実施する。これらアッセイのいずれかにおいて、IGSFP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、または複数の試験化合物をスクリーニングすることができる。

10

【0163】

別の実施態様では、胚性幹細胞（ES細胞）における相同組み換えを用いて動物モデル系内で、IGSFP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号および第5,767,337号などを参照）。例えば129/SvJ細胞系などのマウス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され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。或いは、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マウス株などから採取したマウス細胞胚盤胞に微量注入する。これらの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特定し、これら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。

20

【0164】

IGSFP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、中胚葉および外胚葉の細胞タイプを含む、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、造血系統および心筋細胞に分化する（Thomson, J.A.他 (1998) Science 282:1145-1147）。

30

【0165】

IGSFPをコードするポリヌクレオチドを用いて、ヒト疾患をモデルとした「ノックイン」ヒト化動物（ブタ）または遺伝子組換え動物（マウスまたはラット）を作製することが可能である。ノックイン技術を用いて、IGSFP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試験し、潜在的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IGSFPを乳汁内に分泌するなどIGSFP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）。

40

【0166】

（治療）

IGSFPの領域と免疫グロブリンスーパーファミリータンパク質の領域との間に、例えば配列およびモチーフの内容における化学的類似性および構造的類似性が存在する。更に、IGSFPの発現はCD4⁺ T及び末梢血液細胞、さらに脳、結腸、罹患皮膚、罹患肺、海馬、脾臓、罹患虫部組織と密接に関連する。従って、IGSFPは、免疫系の疾患、神経の疾患、発生または発達障害、筋疾患および細胞増殖異常においてある役割を果たすと考えられる。IGSFP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、IGSFPの発現または活性を低下させることが望ましい。また、IGSFPの発現または活性の低下に関連する疾患の治療

50

においては、IGSFの発現または活性を亢進させることが望ましい。

【0167】

従って、一実施例において、IGSF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、被験者にIGSFまたはその断片や誘導体を投与することが可能である。限定するものではないが、このような疾患として、免疫系の疾患の中には、後天性免疫不全症候群(AIDS)、ブルートン型伴性無ガンマロブリン血症、分類不能型免疫不全(CVI)、ディジョージ症候群(胸腺形成不全症)、胸腺異型性、IgA単独欠損症、重症複合型免疫不全(SCID)、血小板減少および湿疹を伴う免疫不全症(ウイスコット アルドリッチ症候群)、チェディアック 東症候群、慢性肉芽腫症、遺伝性血管神経性浮腫、クッシング病に関連した免疫不全症、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、貧血症、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫性多腺性内分泌カンジダ性外胚葉ジストロフィ(APECED)、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー皮膚炎、皮膚筋炎、真性糖尿病、肺気腫、リンパ球毒素性一時的リンパ球減少症、胎児赤芽球症、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、心筋または心膜炎症、骨関節炎、骨粗しょう症、腓炎、乾癬、ライター症候群、慢性関節リウマチ炎、硬皮症、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、血小板減少症、潰瘍性大腸炎、ウェルナー症候群、癌合併症、血液透析、体外循環、ウイルス感染症、細菌感染症、真菌感染症、寄生虫感染症、原虫感染症、蠕虫感染症、外傷が含まれ、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、プリオン病(クールー、クロイツフェルト ヤコブ病、ゲルストマン症候群、及びGerstmann-Straussler-Scheinker症候群を含む)、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、ダウン症候群を含む中枢神経系の精神薄弱及び他の発生障害、脳性麻痺、神経骨格異常症、自律神経系障害、脳神経障害、脊髄障害、筋ジストロフィー及びその他の神経筋障害、末梢神経疾患、皮膚筋炎及び多発性筋炎、遺伝性、代謝性、内分泌性、及び中毒性の筋疾患、重症筋無力症、周期性四肢麻痺、精神病(気分性、不安性の障害、分裂病性疾患)、季節性障害(SAD)、静座不能、健忘症、緊張病、糖尿病性ニューロパシー、錐体外路性終末欠陥症候群、ジストニー、分裂病性精神障害、帯状疱疹後神経痛、及びトゥレット病と、進行性核上麻痺、大脳皮質基底核部変性症、及び家族性の前頭側頭性健忘症とが含まれ、発生または発達障害の中には尿細管性アシドーシス、貧血、クッシング症候群、軟骨形成不全性小人症、デュシェンヌベッカー型筋ジストロフィー、癲癇、性腺形成異常、WAGR症候群(ウィルムス腫瘍、無虹彩症、尿生殖器異常、精神薄弱)、スミス マジェニス症候群(Smith-Magenis syndrome)、脊髄形成異常症候群、遺伝性粘膜上皮異形成、遺伝性角皮症、シャルコー マリー ツース病及び神経線維腫症などの遺伝性神経病、甲状腺機能低下症、水頭症、Sydenham舞踏病(Sydenham's chorea)及び脳性小児麻痺などの発作障害、脊髄二分裂、無脳症、頭蓋脊椎披裂、先天性緑内障、白内障、感覚神経性聴力損失が含まれ、筋疾患には、心筋症、心筋炎、デュシェンヌ型筋ジストロフィー、ベッカー型筋ジストロフィー、筋緊張(強直)性ジストロフィー、セントラルコア病、ネマリンミオパシー、中心核ミオパシー、脂質ミオパシー、ミトコンドリアミオパシー、感染性筋炎、多発性筋炎、皮膚筋炎、封入体筋炎、甲状腺中毒性ミオパシーおよびエタノールミオパシーが含まれる。細胞増殖異常の中には日光性角化症、動脈硬化、アテローム性動脈硬化、滑液包炎、硬変、肝炎、混合型結合組織病(MCTD)、骨髄線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症、並びに腺癌及び白血病、リンパ腫、黒色腫、骨髄腫、肉腫、及び奇形癌、

10

20

30

40

50

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、子宮の癌が含まれる。

【0168】

別の実施態様では、IGSFP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、上記しだけに限られるものではないが疾患を含むIGSFP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。

【0169】

更に別の実施例では、限定するものではないが上に列記した疾患を含むIGSFPの発現または活性の低下に関連した疾患の治療または予防のために、実質的に精製されたIGSFPを含む組成物を好適な医薬用キャリアと共に患者に投与することも可能である。

10

【0170】

更に別の実施態様では、限定するものではないが上に列記した疾患を含む、IGSFP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、IGSFPの活性を調節するアゴニストを被験者に投与し得る。

【0171】

更なる実施態様では、IGSFPの発現または活性の増大に関連した疾患の治療または予防のために、被験者にIGSFPのアнтаゴニストを投与し得る。限定するものではないが、このような疾患の例には、上記した免疫系の疾患、神経の疾患、発生または発達障害、筋疾患および細胞増殖異常が含まれる。一実施態様では、IGSFPと特異的に結合する抗体が直接アンタゴニストとして、或いはIGSFPを発現する細胞または組織に薬剤を運ぶターゲティング或いは送達機構として間接的に用いられ得る。

20

【0172】

別の実施態様では、限定するものではないが上に列記した疾患を含む、IGSFPの発現または活性の増大に関連した疾患の治療または予防のために、IGSFPをコードするポリヌクレオチドの相補配列を発現するベクターを被験者に投与し得る。

【0173】

別の実施態様では、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ることもできる。併用療法で用いる好適な治療薬は、当業者が従来の医薬原理に従って選択し得る。治療薬と組み合わせることにより、上記した種々の疾患の治療または予防に相乗効果をもたらし得る。この方法により、少量の各薬物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。

30

【0174】

IGSFPのアнтаゴニストは、当分野で一般的な方法を用いて製造できる。特に精製されたIGSFPを用いて抗体を作ったり、治療薬のライブラリ群をスクリーニングしてIGSFPと特異結合する薬物を同定できる。IGSFPへの抗体も、当分野で周知の方法を用いて産生され得る。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、Fab断片、およびFab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(すなわ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。一本鎖抗体(例えば、ラクダまたはラマ)は有力な酵素阻害剤であり、またペプチド擬態物質の設計および免疫吸着剤やバイオセンサーの開発に利点があるであろう(Muylderms, S. (2001) J. Biotechnol. 74:277-302)。

40

【0175】

抗体の産生のためには、ヤギ、ウサギ、ラット、マウス、ラクダ、ヒトコブラクダ、ラマ、ヒトなどを含む種々の宿主が、IGSFP、若しくは免疫原性の特性を備えるその任意の断片またはオリゴペプチドの注入によって免疫化され得る。宿主の種に応じて、種々のアジュバントを用いて免疫応答を高めることもできる。限定するものではないがこのようなアジュバントには、フロイントアジュバントと、水酸化アルミニウムなど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

50

剤、スカシガイのヘモシアニン (KLH)、ジニトロフェノールなどの界面活性剤とがある。ヒトに用いられるアジュバントの中では、BCG (カルメット グラン杆菌) 及びコリネバクテリウム パルバム (*Corynebacterium parvum*) が特に好ましい。

【0176】

IGSFPI に対する抗体を誘発するために用いられるオリゴペプチド、ペプチド、または断片は、少なくとも約5個のアミノ酸からなり、一般的には約10個以上のアミノ酸からなるものが好ましい。これらのオリゴペプチド、ペプチドまたは断片は、天然のタンパク質のアミノ酸配列の一部と同一であることも望ましい。IGSFPI アミノ酸類の短いストレッチ群は、KLHなど他のタンパク質の配列と融合され、このキメラ分子に対する抗体群が産生され得る。

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

【0177】

更に、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「キメラ抗体」作製のために開発した技術が、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 (Morrison, S.L. 他 (1984) *Proc. Natl. Acad. Sci. USA* 81:6851-6855, Neuberger, M.S. 他 (1984) *Nature* 312:604-608, Takeda, S. 他 (1985) *Nature* 314:452-454等を参照)。別法では、当分野で周知の方法を用いて、一本鎖抗体の産生のための記載された技術を適用して、IGSFPI 特異的な一本鎖抗体を生成する。関連した特異性を有するがイデ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類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)。

【0178】

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、或いは文献に開示されているように非常に特異的な結合試薬の免疫グロブリンのライブラリまたはパネルのスクリーニングによっても行い得る (Orlandi, R. 他 (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837, Winter, G. 他 (1991) *Nature* 349:293-299等を参照)。

【0179】

IGSFPI に対する特異的な結合部位を含む抗体断片も得ることができる。例えば、限定するものではないが、このような断片には、抗体分子のペプシン消化によって作製される $F(ab')_2$ 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、所望の特異性を持つモノクローナル Fab 断片を迅速且つ容易に同定することが可能となる (Huse, W.D. 他 (1989) *Science* 246:1275-1281等を参照)。

【0180】

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競合結合試験、または免疫放射定量測定法のための数々のプロトコルが、当分野では周知である。通常このようなイムノアッセイには、IGSFPI とその特異性抗体との間の複合体形成の計測が含まれる。2つの非干渉性 IGSFPI エピトープに対して反応性を持つモノクローナル抗体群を用いる、2部位モノクローナルベースのイムノアッセイが一般に利用されるが、競合結合試験も利用できる (Pound, 前出)。

【0181】

ラジオイムノアッセイ技術と共に Scatchard 分析などの様々な方法を用いて、IGSFPI に対する抗体の親和性を評価する。親和性を結合定数 K_a で表すが、この K_a は、平衡状態の下で IGSFPI 抗体複合体のモル濃度を遊離抗体と遊離抗原のモル濃度で除して得られる値である。

10

20

30

40

50

多数のIGSFPエピトープに対して親和性が不均一なポリクローナル抗体医薬の K_a は、IGSFPに対する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或る特定のIGSFPエピトープに単一特異的であり、或るモノクローナル抗体製剤について判定した K_a は、親和性の真の測定値を表す。 K_a 値が $10^9 \sim 10^{12}$ liter/molの高親和性抗体試薬は、IGSFP抗体複合体が過酷な処理に耐えなければならないイムノアッセイに用いるのが好ましい。 K_a 値が $10^6 \sim 10^7$ liter/molの低親和性抗体医薬は、IGSFPが抗体から最終的に活性化状態で解離する必要がある免疫精製(immunopurification)及び類似の処理に用いるのが好ましい(Catty, D. (1988) Antibodies, Volume I: A Practical Approach. IRL Press, Washington, DC、Liddell, J. E.およびCryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY)。

10

【0182】

ポリクローナル抗体製剤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製剤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1~2mg/mlの特異的な抗体、好ましくは5~10mg/mlの特異的な抗体を含むポリクローナル抗体製剤は一般に、IGSFP抗体複合体を沈殿させなければならない処理に用いられる。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他の文献等を参照)。

【0183】

本発明の別の実施例では、IGSFPをコードするポリヌクレオチド、またはその任意の断片や相補配列が、治療目的で使用されうる。ある実施態様では、IGSFP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、RNA、PNAあるいは修飾オリゴヌクレオチド)を設計して遺伝子発現を変更することができる。このような技術は当分野では周知であり、アンチセンスオリゴヌクレオチドまたは大きな断片が、IGSFPをコードする配列の制御領域から、またはコード領域に沿ったさまざまな位置から設計可能である。(Agrawal, S.編集(1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ等を参照)。

20

【0184】

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作製する発現プラスミドの形で細胞内に送達することが可能である(例えばSlater, J.E.他(1998) J. Allergy Clin. Immunol. 102(3):469-475、Scanlon, K.J.他(1995) 9(13):1288-1296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随伴ウイルスベクターなどのウイルスベクターを用いて細胞内に導入することもできる(Miller, A.D. (1990) Blood 76:271、前出のAusubel、Uckert, W.及びW. Walther (1994) Pharmacol. Ther. 63(3):323-347等を参照)。その他の遺伝子送達機構には、リポソーム系、人工的なウイルスエンベロープおよび当分野で公知のその他のシステムが含まれる(例えばRossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J.他(1998) J. Pharm. Sci. 87(11):1308-1315、Morris, M.C.他(1997) Nucleic Acids Res. 25(14):2730-2736を参照)。

30

40

【0185】

本発明の別の実施例では、IGSFPをコードするポリヌクレオチドを、体細胞若しくは生殖細胞の遺伝子治療に用いることが可能である。遺伝子治療を行うことにより、(i) 遺伝子欠損症(例えばX染色体鎖遺伝(Cavazzana-Calvo, M.他(2000) Science 288:669-672)により特徴付けられる重度の複合型免疫欠損(SCID)-X1の場合)、先天性アデノシンデアミナーゼ(ADA)欠損症に関連する重度の複合型免疫欠損(Blaese, R.M.他(1995) Science 270:475-480、Bordignon, C.他(1995) Science 270:470-475)、嚢胞性繊維症(Zabner, J.他(1993) Cell 75:207-216; Crystal, R.G.他(1995) Hum. Gene Therapy 6:643-666、Crystal, R.G.他(1995) Hum. Gene Therapy 6:667-703)、サラセミア(thalassamia)、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

50

る血友病 (Crystal, R.G. (1995) Science 270:404-410、Verma, I.M. 及び Somia. N. (1997) Nature 389:239-242) を治療し、(ii) 条件的致死性遺伝子産物を発現させ (例えば制御不能な細胞増殖に起因する癌の場合)、(iii) 細胞内の寄生虫 (例えばヒト免疫不全ウイルス (HIV) (Baltimore, D. (1988) Nature 335:395-396、Poesebla, 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等の真菌寄生虫、並びに Plasmodium falciparum及びTrypanosoma cruzi等の原虫寄生体に対する防御機能を有するタンパク質を発現させることができる。IGSFPの発現若しくは調節に必要な遺伝子の欠損が疾患を引き起こす場合、導入した細胞の好適な集団からIGSFPを発現させて、遺伝子欠損によって起こる症状の発現を緩和することが可能である。

10

【0186】

本発明の更なる実施例では、IGSFPの欠損による疾患や異常症は、IGSFPをコードする哺乳動物発現ベクターを作製して、これらのベクターを機械的手段によってIGSFP欠損細胞に導入することによって治療する。in vivo或いはex vitroの細胞に用いる機械的導入技術には、(i) 個々の細胞内への直接的なDNA微量注射法、(ii) 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。

IGSFPの発現に影響を及ぼし得る発現ベクターには、限定するものではないが、PCDNA 3.1、EPITAG、PRCCMV2、PREP、PVAX、PCR2-TOPOTAベクター (Invitrogen, Carlsbad CA)、PCM V-SCRIPT、PCMV-TAG、PEGSH/PERV (Stratagene, La Jolla CA)、PTET-OFF、PTET-ON、PTRE2、PTRE2-LUC、PTK-HYG (Clontech, Palo Alto CA) が含まれる。IGSFPを発現させるために、(i) 恒常的に活性なプロモーター (例えば、サイトメガロウイルス (CMV)、ラウス肉腫ウイルス (RSV)、SV40ウイルス、チミジンキナーゼ (TK)、若しくは - アクチン遺伝子等)、(ii) 誘導性プロモーター (例えば、市販されているT-REXプラスミド (Invitrogen) に含まれている、テトラサイクリン調節性プロモーター (Gossen, M. 及び H. Bujard (1992) Proc. Natl. Acad. Sci. U.S.A. 89:5547-5551、Gossen, M. 他 (1995) Science 268:1766-1769; Rossi, F.M.V. 及び H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456))、エクジソン誘導性プロモーター (市販されているプラスミドPVGRXR及びPINDに含まれている: Invitrogen)、FK506/ラパマイシン誘導性プロモーター、またはRU486/ミフェプリストーン誘導性プロモーター (Rossi, F.M.V. 及び H.M. Blau, 前出)、または (iii) 正常な個体に由来するIGSFPをコードする内在性遺伝子の天然のプロモーター若しくは組織特異的プロモーターを用いることが可能である。

20

30

【0187】

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、当業者はポリヌクレオチドを培養中の標的細胞に送達することが可能になる。また、実験パラメータ群を最適化するのに必要な努力が最小限になる。別法では、リン酸カルシウム法 (Graham, F.L. 及び A.J. Eb (1973) Virology 52:456-467) 若しくは電気穿孔法 (Neumann, B. 他 (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、標準化された哺乳類の形質移入プロトコルの修正が必要である。

40

【0188】

本発明の別の実施例では、IGSFPの発現に関連する遺伝子欠損によって起こる疾患や異常症は、(i) レトロウイルス末端反復配列 (LTR) プロモーター若しくは独立したプロモーターのコントロール下でIGSFPをコードするポリヌクレオチドと、(ii) 好適なRNAパッケージングシグナルと、(iii) 追加のレトロウイルス・シス作用性RNA配列及び効率的なベクターの増殖に必要なコーディング配列を伴うRev応答性エレメント (RRE) とからなるレトロウイルスベクターを作製して治療することができる。レトロウイルスベクター (例えばPFB及びPFBNE0) はStratagene社から市販されており、刊行データ (Riviere, I. 他. (

50

1995) Proc. Natl. Acad. Sci. U.S.A. 92:6733-6737) に基づいている。上記データを引用することをもって本明細書の一部とする。このベクターは、好適なベクター産生細胞株 (VPCL) において増殖される。VPCLは、各標的細胞上の受容体への親和性を持つエンベロープ遺伝子を、またはVSVgなど汎親和性エンベロープタンパク質を発現する (Armentano, D. 他 (1987) J. Virol. 61:1647-1650、Bender, M.A. 他 (1987) J. Virol. 61:1639-1646、Adam, M.A. および A.D. Miller (1988) J. Virol. 62:3802-3806、Dull, T. 他 (1998) J. Virol. 72:8463-8471、Zufferey, R. 他 (1998) J. Virol. 72:9873-9880)。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株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. 他 (1997) J. Virol. 71:7020-7029、Bauer, G. 他 (1997) Blood 89:2259-2267、Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716、Ranga, U. 他 (1998) Proc. Natl. Acad. Sci. U.S.A. 95:1201-1206、Su, L. (1997) Blood 89:2283-2290)。

10

【0189】

別法では、アデノウイルス系遺伝子治療の送達系を用いて、IGSFPの発現に関連する1つ以上の遺伝子異常を有する細胞にIGSFPをコードするポリヌクレオチドを送達する。アデノウイルス系ベクター類の作製およびパッケージングについては、当業者に周知である。複製欠損型アデノウイルスベクター類は、種々の免疫調節タンパク質をコードする遺伝子群を、無損傷の臍島内に導入する目的で多様に利用し得ることが証明された (Csete, M.E. 他 (1995) Transplantation 27:263268)。使用できる可能性のあるアデノウイルスベクターは、Armentanoに付与された米国特許第5,707,618号 (「Adenovirus vectors for gene therapy」) に記載されており、引用することをもって本明細書の一部とする。アデノウイルスベクターについては、Antinozzi, P.A. 他 (1999) Annu. Rev. Nutr. 19:511-544、Verma, I.M. 及び 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。

20

【0190】

別法では、ヘルペス系遺伝子治療の送達系を用いて、IGSFPの発現に関連する1つ以上の遺伝子異常を有する標的細胞に、IGSFPをコードするポリヌクレオチドを送達する。単純疱疹ウイルス (HSV) 系のベクターは、HSV親和性の中枢神経細胞にIGSFPを導入する際に特に有用たりうる。ヘルペス系ベクター類の作製およびパッケージングは、当業者に公知である。或る複製適格性単純ヘルペスウイルス (HSV) 1型系のベクターが、或るレポーター遺伝子の、霊長類の眼への送達に用いられている (Liu, X. 他 (1999) Exp. Eye Res. 169:385395)。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 strains 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の使用についての記載がある。上記特許はまた、ICP4、ICP27及びICP22を欠失した組換え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. 他 (1999) J. Virol. 73:519-532及び Xu, H. 他 (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セグメント群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。

30

40

【0191】

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IGSFPをコードするポリヌクレオチドを標的細胞に送達する。プロトタイプのウイルスであるセムリキ森林熱

50

ウイルス (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はSFVゲノムに基づいている (Garoff, H. 及び K.-J. Li (1998) *Cun. Opin. Biotech.* 9:464-469)。

ウイルスRNAの複製中に、通常はウイルスのカプシドタンパク質をコードするサブゲノムRNAが作り出される。このサブゲノムRNAは、完全長のゲノムRNAより高いレベルに複製されるため、酵素活性 (例えばプロテアーゼ及びポリメラーゼ) を有するウイルスタンパク質に比べてカプシドタンパク質が過剰産生される。同様に、IGSFPをコードする配列を

ウイルスゲノムのカプシドをコードする領域に導入することによって、ベクター導入細胞において多数のIGSFPをコードするRNAが産生され、高いレベルでIGSFPが合成される。

通常、ウイルスの感染は、数日以内の細胞溶解を伴う。一方、シンドビスウイルス (SIN) の或る変異体を有するハムスター正常腎臓細胞 (BHK-21) 群が持続的な感染を確立する能力は、ウイルス類の溶解複製を、遺伝子治療に応用し得るように好適に改変可能であることを示唆する (Dryga, S.A. 他 (1997) *Virology* 228:74-83)。ウイルスの宿主域は広いので、様々なタイプの細胞にIGSFPを導入することができる。或る集団における或るサブセットの細胞群の特異的形質導入は、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およびRNAの形質移入方法およびウイルスの感染方法は、当業者に公知である。

【0192】

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 (transcription initiation site) とは例えばスタート部位 (start site) 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。三重らせん体塩基対形成は、二重らせん体が開いてポリメラーゼ、転写因子または調節分子と結合するのを阻害するので有用である。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. 他 (1994) in: Huber, B.E. and B.I. Carr, *Molecular and Immunologic Approaches*, Futura Publishing Co., Mt. Kisco, NY, 163-177ページ等を参照)。相補配列またはアンチセンス分子もまた、転写物がリボソームに結合するのを阻止することによってmRNAの翻訳を阻止するべく設計することができる。

【0193】

リボザイムは酵素的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的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与している。例えば、人工的に作製されたハンマーヘッド型リボザイム分子が、IGSFP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できる可能性がある。

【0194】

任意の潜在的RNA標的内の特異的リボザイム切断部位は、GUA、GUU、GUC配列を含めたりボザイム切断部位に対する標的分子をスキャンすることによって先ず同定される。一度同定されると、切断部位を含む標的遺伝子の領域に対応する15~20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群とのハイブリダイゼーションへのアクセス可能性をテストすることによって行い得る。

【0195】

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既知の任意の方法を用いて作製し得る。作製方法には、固相フォスフォアミダイト化学合成等のオリゴヌクレオチドを化学的に合成する方法がある。或いは、IGSFP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などの好適なRNAポリメラーゼプロモーターを用いて多様なベクター内に取り込むことが可能である。或いは、相補的RNAを構成的或いは誘導的に合成するようなこれらcDNA産物を、細胞株、細胞または組織内に導入することができる。

【0196】

10

20

30

40

50

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としては、分子の5'末端、3'末端、あるいはその両方において隣接配列群を追加することや、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オチオエートまたは2' O-メチルを使用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るが、これら全ての分子に拡大することができる。そのためには、内因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や、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) 等を含める。

【0197】

本発明の更なる実施例は、IGSFPをコードするポリヌクレオチドの発現の変化に有効な化合物をスクリーニングする方法を含む。限定するものではないが特異ポリヌクレオチドの発現変化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、IGSFP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、IGSFP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、IGSFP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、IGSFP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。

【0198】

特異ポリヌクレオチドの発現改変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、既に市販のまたは私的な、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。IGSFPをコードするポリヌクレオチドを含むサンプルは、このようにして得られた試験化合物の少なくとも1つに曝露する。サンプルには例えば、無傷細胞、または透過化処理した細胞、あるいは *in vitro* 無細胞系すなわち再構成生化学系があり得る。IGSFP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、IGSFP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改変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発現改変に有効な化合物に対して、例えば分裂酵母 (*Schizosaccharomyces pombe*) 遺伝子発現系 (Atkins, D. 他 (1999) 米国特許第5,932,435号、Arndt, G.M. 他 (2000) *Nucleic Acids Res.* 28:E15) またはHeLa細胞等のヒト細胞系 (Clarke, M.L. 他 (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等) 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. 他 (1997) 米国特許第5,686,242号、Bruice, T.W. 他 (2000) 米国特許第6,022,691号)。

【0199】

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

10

20

30

40

50

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による、またはリボソーム注入や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（例えば Goldman, C.K.他 (1997) Nat. Biotechnol. 15:462-466を参照）。

【0200】

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。

【0201】

本発明のさらなる実施態様は、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組成物の投与に関連する。賦形剤には例えば、糖、でんぷん、セルロース、ゴムおよびタンパク質がある。様々な剤型が広く知られており、詳細はRemington's Pharmaceutical Sciences (Maack Publishing, Easton PA)の最新版に記載されている。このような組成物は、IGSFP、IGSFPの抗体、擬態、アゴニスト、アンタゴニスト、またはIGSFPのインヒビターなどからなる。

10

【0202】

本発明に用いられる組成物は、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。

【0203】

肺から投与する組成物は、液状または乾燥粉末状で調製し得る。このような組成物は通常、患者が吸入する直前にエアロゾル化する。小分子（例えば伝統的な低分子量有機薬）の場合には、速効製剤のエアロゾル送達は当分野で公知である。高分子（例えばより大きなペプチド及びタンパク質）の場合には、当該分野において肺の肺泡領域を介しての肺送達が最近向上したことにより、インスリン等の薬剤を実質的に血液循環へ輸送することを可能にした（Patton, J.S. 他, 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。

20

【0204】

本発明での使用に適した組成物には、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組成物が含まれる。有効投与量の決定は、当業者の能力の範囲内で行う。IGSFPを有する、またはその断片群を有する高分子群を直接、細胞内に送達すべく、特殊な種々の形状の組成物群が調製され得る。例えば、細胞不透過性高分子を含むリボソーム製剤は、その高分子の細胞融合と細胞内送達とを促進し得る。別法では、IGSFPまたはその断片をHIV Tat-1タンパク質の陽イオン性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類は、或るマウスモデル系の、脳を含む全ての組織の細胞群に形質導入することがわかっている（Schwarze, S.R. 他 (1999) Science 285:1569-1572）。

30

【0205】

任意の化合物に対して、細胞培養アッセイ、例えば新生物性細胞の細胞培養アッセイにおいて、或いは、動物モデル、例えばマウス、ラット、ウサギ、イヌ、サルまたはブタ等において、先ず治療上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、好適な濃度範囲およ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および投与経路を決定することができる。

40

【0206】

治療有効投与量とは、症状や容態を回復させる、たとえばIGSFPまたはその断片、IGSFPの抗体、IGSFPのアゴニストまたはアンタゴニスト、インヒビターなどの活性成分の量を指す。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%の治療有効量）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。毒性効果の治療効果に対する投与量比が治療指数であり、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組成物が望ましい。細胞培養アッセイ及び動物実験から得られたデータは、ヒトに用いるための投与量の範囲

50

を策定するのに用いられる。このような組成物に含まれる投与量は、毒性を殆どあるいは全く持たず、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、患者の感受性およ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。

【0207】

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。十分なレベルの活性成分を与え、あるいは所望の効果を維持すべく、用法および用量を調整する。被験者に関する要素としては、疾患の重症度、被験者の全身健康状態、被験者の年齢、体重及び性別(ジェンダー)、投与の時間及び頻度、薬剤の併用、反応感受性及び治療に対する応答等を考慮しうる。作用期間が長い組成物は、特定の製剤の半減期及びクリアランス率によって、3~4日毎に1度、1週間に1度、或いは2週間に1度の間隔で投与し得る。

10

【0208】

通常の投与量は、投与の経路にもよるが約0.1~100.000 µgであり、合計で約1gまでとする。特定の投与量及び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、症状、部位などに特異的なものとなる。

【0209】

(診断)

別の実施例では、IGSFPIに特異的に結合する抗体が、IGSFPIの発現によって特徴付けられる疾患の診断、またはIGSFPIやIGSFPIのアゴニストまたはアンタゴニスト、インヒビターで治療を受けている患者をモニターするためのアッセイに用いられる。診断目的に有用な抗体は、上記の治療の箇所で記載した方法と同じ方法で調合される。IGSFPIの診断アッセイには、ヒトの体液から、あるいは細胞や組織の抽出物から、抗体および標識を用いてIGSFPI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。

20

【0210】

IGSFPIを測定するためのELISA, RIA, 及びFACSを含む種々のプロトコルは、当分野では周知であり、変容した或いは異常なレベルのIGSFPIの発現を診断するための基盤を提供する。正常或いは標準的なIGSFPIの発現の値は、複合体の形成に適した条件下で、正常な哺乳動物、例えばヒトなどの被験者から採取した体液または細胞とIGSFPIに対する抗体とを結合させることによって決定する。標準複合体形成量は、種々の方法、例えば測光法で定量できる。被験者、対照、及び疾患生検組織からの各サンプルのIGSFPIの発現の量が基準値と比較される。標準値と被験体との偏差が、疾患を診断するパラメータを確定する。

30

【0211】

本発明の別の実施例では、IGSFPIをコードするポリヌクレオチドを診断のために用いることもできる。用いることができるポリヌクレオチドには、オリゴヌクレオチド配列、相補的RNA及びDNA分子、そしてPNAが含まれる。このポリヌクレオチドを用いて、疾患と相関し得るIGSFPIを発現する生検組織における遺伝子の発現を検出し定量する。この診断アッセイを用いて、IGSFPIの存在の有無、更に過剰な発現を調べ、治療中のIGSFPI値の調節を監視する。

40

【0212】

一実施形態では、IGSFPIまたは近縁の分子をコードする遺伝子配列を含むポリヌクレオチド配列を検出可能なPCRプローブを用いたハイブリダイゼーションによって、IGSFPIをコードする核酸配列を同定することが可能である。例えば5'調節領域である高度に特異的な領域か、例えば保存されたモチーフであるやや特異性の低い領域から作られているかのプローブの特異性と、ハイブリダイゼーション或いは増幅のストリンジェントは、プローブ

50

がIGSFPをコードする自然界の配列のみを同定するかどうか、或いは対立遺伝子や関連配列コードする自然界の配列のみを同定するかどうかによって決まるであろう。

【0213】

プローブはまた、関連する配列の検出に利用され、IGSFPをコードする任意の配列と少なくとも50%の配列同一性を有し得る。目的の本発明のハイブリダイゼーションプローブには、DNAあるいはRNAが可能であり、SEQ ID NO:13-24の配列、或いはIGSFP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。

【0214】

IGSFPをコードするDNAに対して特異的なハイブリダイゼーションプローブの作製方法には、IGSFP及びIGSFP誘導体をコードするポリヌクレオチド配列をmRNAプローブの作製のためのベクターにクローニングする方法がある。mRNAプローブ作製のためのベクターは、当業者に知られており、市販されており、好適なRNAポリメラーゼ及び好適な標識されたヌクレオチドを加えることによって、*in vitro*でRNAプローブを合成するために用いられ得る。ハイブリダイゼーションプローブは、種々のレポーターの集団によって標識され得る。レポーター集団の例としては、³²Pまたは³⁵S等の放射性核種、或いはアビジン/ビオチン結合系を介してプローブに結合されたアルカリホスファターゼ等の酵素標識などが挙げられる。

【0215】

IGSFPをコードするポリヌクレオチド配列群を用いて、IGSFPの発現に関連する疾患を診断できる。限定するものではないが、このような疾患としては、後天性免疫不全症候群(AIDS)、ブルトン型伴性無ガンマロブリン血症、分類不能型免疫不全(CVI)、ディジョージ症候群(胸腺形成不全症)、胸腺異型性、IgA単独欠損症、重症複合型免疫不全(SCID)、血小板減少および湿疹を伴う免疫不全症(ウイスコット アルドリッチ症候群)、チェディアック 東症候群、慢性肉芽腫症、遺伝性血管神経性浮腫、クッシング病に関連した免疫不全症、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、貧血症、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫性多腺性内分泌カンジダ性外胚葉ジストロフィ(APECED)、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー皮膚炎、皮膚筋炎、真性糖尿病、肺気腫、リンパ球毒素性一時性リンパ球減少症、胎児赤芽球症、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、腭炎、乾癬、ライター症候群、慢性関節リウマチ炎、硬皮症、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、血小板減少症、潰瘍性大腸炎、ウェルナー症候群、癌合併症、血液透析、体外循環、ウイルス感染症、細菌感染症、真菌感染症、寄生虫感染症、原虫感染症、蠕虫感染症、外傷が含まれ、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、プリオン病(クールー、クロイツフェルト ヤコブ病、ゲルストマン症候群、及びGerstmann-Straussler-Scheinker症候群を含む)、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、ダウン症候群を含む中枢神経系の精神薄弱及び他の発生障害、脳性麻痺、神経骨格異常症、自律神経系障害、脳神経障害、脊髄障害、筋ジストロフィー及びその他の神経筋障害、末梢神経疾患、皮膚筋炎及び多発性筋炎、遺伝性、代謝性、内分泌性、及び中毒性の筋疾患、重症筋無力症、周期性四肢麻痺、精神病(気分性、不安性の障害、分裂病性疾患)、季節性障害(SAD)、静座不能、健忘症、緊張病、糖尿病性ニューロパシー、錐体外路性終末欠陥症候群、ジストニー、分裂病性精神障害、帯状疱疹後神経痛、及びトウレット病と、進行性核上麻痺、大脳

10

20

30

40

50

皮質基底核部変性症、及び家族性の前頭側頭性健忘症とが含まれ、発生または発達障害の中には尿細管性アシドーシス、貧血、クッシング症候群、軟骨形成不全性小人症、デュシェンヌ型筋ジストロフィー、癲癇、性腺形成異常、WAGR症候群（ウィルムス腫瘍、無虹彩症、尿生殖器異常、精神薄弱）、スミス マジェニス症候群（Smith-Magenis syndrome）、脊髄形成異常症候群、遺伝性粘膜上皮異形成、遺伝性角皮症、シャルコーマリー ツース病及び神経線維腫症などの遺伝性神経病、甲状腺機能低下症、水頭症、Syndenham舞蹈病（Syndenham's chorea）及び脳性小児麻痺などの発作障害、脊髄二分裂、無脳症、頭蓋脊椎披裂、先天性緑内障、白内障、感覚神経性聴力損失が含まれ、筋疾患には、心筋症、心筋炎、デュシェンヌ型筋ジストロフィー、ベッカー型筋ジストロフィー、筋緊張（強直）性ジストロフィー、セントラルコア病、ネマリンミオパシー、中心核ミオパシー、脂質ミオパシー、ミトコンドリアミオパシー、感染性筋炎、多発性筋炎、皮膚筋炎、封入体筋炎、甲状腺中毒性ミオパシーおよびエタノールミオパシーが含まれる。細胞増殖異常の中には日光性角化症、動脈硬化、アテローム性動脈硬化、滑液包炎、硬変、肝炎、混合型結合組織病（MCTD）、骨髄線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症、並びに腺癌及び白血病、リンパ腫、黒色腫、骨髄腫、肉腫、及び奇形癌、具体的には、副腎、膀胱、骨、骨髄、脳、乳房、頸部、胆嚢、神経節、消化管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、子宮の癌が含まれる。IGSFをコードするポリヌクレオチド配列は、変容したIGSF発現を検出するために患者から採取した体液或いは組織を利用する、サザン法やノーザン法、ドットプロット法、或いはその他の膜系の技術、PCR法や

10

20

30

40

50

【0216】

ある実施態様では、IGSFをコードするヌクレオチド配列は、関連する疾患、特に上記した疾患を検出するアッセイにおいて有用であろう。IGSFをコードするヌクレオチド配列を標準方法で標識化し、ハイブリダイゼーション複合体の形成に好適な条件下で、患者から採取した体液或いは組織のサンプルに加えることができる。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。患者サンプル中のシグナルの量が、対照サンプルと較べて著しく変化している場合は、該サンプル内の、IGSFをコードするヌクレオチド配列群のレベル変化の存在が、関連する障害の存在を標示する。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、ある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。

【0217】

IGSFの発現に関連する疾患の診断の基準となるものを提供するために、発現の正常すなわち標準的なプロファイルが確立される。これは、ハイブリダイゼーション或いは増幅に好適な条件の下で、動物或いはヒトの何れかの正常な被験者から採取された体液或いは細胞抽出物と、IGSFをコードする配列或いはその断片とを混合することにより達成される。実質的に精製されたポリヌクレオチドを既知量で用いて行った実験から得た値を正常な被験者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。

【0218】

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定期的に繰り返し得る。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。

【0219】

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

方法を提供し得る。この種のより明確な診断により、医療の専門家が予防方法または積極的な治療法を早くから利用し、それによって癌の発生または更なる進行を防止することが可能となる。

【0220】

IGSFPをコードする配列から設計されたオリゴヌクレオチドのさらなる診断への利用には、PCRの利用が含まれ得る。これらのオリゴマーは、化学的に合成するか、酵素により生産するか、あるいは*in vitro*で産出し得る。オリゴマーは、好ましくはIGSFPをコードするポリヌクレオチドの断片、或いはIGSFPをコードするポリヌクレオチドと相補的なポリヌクレオチドの断片を含み、最適な条件の下、特定の遺伝子や条件を識別するために利用される。また、オリゴマーは、やや緩いストリンジェント条件下で、近縁のDNA或いはRNA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

【0221】

或る実施態様において、IGSFPをコードするポリヌクレオチド配列由来のオリゴヌクレオチドプライマーを用いて、一塩基多型性(SNP)を検出し得る。SNPは、多くの場合にヒトの先天性または後天性遺伝病の原因となるような置換、挿入及び欠失である。限定するものではないがSNPの検出方法には、SSCP(single-stranded conformation polymorphism)及び蛍光SSCP(fSSCP)法がある。SSCPでは、IGSFPをコードするポリヌクレオチド配列由来のオリゴヌクレオチドプライマーを用いたポリメラーゼ連鎖反応(PCR)でDNAを増幅する。DNAは例えば、病変組織または正常組織、生検サンプル、体液その他に由来し得る。DNA内のSNPは、一本鎖形状のPCR生成物の2次及び3次構造に差異を生じさせる。差異は非変性ゲル中でのゲル電気泳動法を用いて検出可能である。fSSCPでは、オリゴヌクレオチドプライマーを蛍光性に標識する。それによってDNAシーケ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SNP(*in silico* SNP, isSNP)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へ構築されるような個々のオーバーラップするDNA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、DNAの実験室での調製に、また統計モデル及びDNA配列クロマトグラムの自動分析を用いたシーケンシングのエラーに起因する配列の変異をフィルタリングして除去する。別の態様では、例えば高処理MASSARRAYシステム(Sequenom, Inc., San Diego CA)を用いた質量分析によりSNPを検出し、特徴付ける。

20

30

【0222】

ヒト疾患の遺伝的根拠を研究するためにSNPを用いることができる。例えば、少なくとも16の一般的SNPが、非インスリン依存型真性糖尿病と関連がある。SNPは、嚢胞性線維症、鎌状赤血球貧血あるいは慢性肉芽腫症等の単一遺伝子病の転帰の差異を研究する面でも有用である。例えば、マンノース結合レクチンの変異体であるMBL2は、嚢胞性線維症の肺での有害な転帰と関連することが示されてきた。SNPはまた、生命を脅かす毒性等の薬剤への患者の反応に影響する遺伝変異体の同定という薬理ゲノミクスにおいても有用性がある。例えば、ALOX5遺伝子のコア・プロモーターにおける或る変異は5-リポキシゲナーゼ経路を標的とする抗喘息剤を用いた治療への臨床反応の減少につながるが、Nアセチルトランスフェラーゼの或る変異は抗結核薬剤イソニアジドに反応する末梢神経障害の高発生率と関連する。異なる集団におけるSNP分布の分析は、集団の起源と移動の追跡以外にも遺伝的浮動、突然変異、組換え、選択の調査において有用である(Taylor, J.G. 他 (2001) Trends Mol. Med. 7:507-512、Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543、Nowotny, P. 他 (2001) Curr. Opin. Neurobiol. 11:637-641)。

40

【0223】

IGSFP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 159:235-244、Duplaa, C. 他 (1993) Anal. Biochem. 212:229-236を参照)。目的のオリゴマーまたはポリヌクレオチドが種々の希釈液中に存在し、分光光度法または比色反応によって定量が迅速になるような

50

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。

【0224】

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物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、突然変異およ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物の活性を開発およ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、有効な治療法を選択するために、この情報を用いて患者の薬理ゲノムプロファイルを開発することができる。例えば、患者の薬理ゲノムプロファイルに基づき、患者に対して高度に効果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。

10

【0225】

別の実施例では、IGSFP、IGSFPの断片、IGSFPに特異的な抗体をマイクロアレイ上のエレメントとして用いることができる。マイクロアレイを用いて、上記のようなタンパク質-タンパク質相互作用、薬物-標的相互作用および遺伝子発現プロファイルをモニターまたは測定することが可能である。

【0226】

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る遺伝子発現の包括的パターンを表す。包括的遺伝子発現パターンは、所与の条件下で所与の時間に発現した遺伝子の数および相対存在量を定量することにより分析し得る (Seilhame r 他 米国特許第5,840,484号「Comparative Gene Transcript Analysis」を参照。当該文献は特に引用することを以って本明細書の一部となす)。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相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相補体がマイクロアレイ上のエレメントのサブセットを複数含むような高処理フォーマットでハイブリダイゼーションを発生させる。結果として得られる転写イメージは、遺伝子活性のプロファイルを提供し得る。

20

30

【0227】

転写イメージは、組織、株化細胞、生検またはその他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。転写イメージはしたがって、組織または生検サンプルの場合には *in vivo*、細胞株の場合には *in vitro* での遺伝子発現を反映する。

【0228】

本発明のポリヌクレオチドの発現のプロファイルを作製する転写物イメージはまた、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示す、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 (toxicant signatures) 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. 他 (1999) Mol. Carcinog. 24 :15 3-159, Steiner, S. 及び N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471、該文献は特に引用することを以って本明細書の一部となす)。試験化合物が既知の毒性を有する化合物のシグネチャと類似のシグネチャを有する場合には、毒性特性を共有している可能性がある。フィンガープリントまたはシグネチャは、多数の遺伝子及び遺伝子ファミリーからの発現情報を含んでいる場合には、最も有用且つ正確である。理想的には、発現をゲノム全域にわたって測定することによって最高品質のシグネチャが得られる。たとえば、発現が任意の試験された化合物によって変容しない遺伝子があったとしても、それらの発現レベルを残りの発現データをノーマライズするために使用できるため、それらの遺伝子は重要である。ノーマライズ手順は、異なる化合物で処理した後の発現データの比較

40

50

に有用である。毒物シグネチャの要素に遺伝子機能を割り当てることで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過程に、遺伝子機能の知識は必要とされない（例えば2000年2月29日に米国国立環境健康科学研究所（National Institute of Environmental Health Sciences）より発行されたPress Release 00-02を参照されたい。これについては<http://www.niehs.nih.gov/oc/news/toxchip.htm>で入手可能である）。従って、毒性シグネチャを用いる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。

【0229】

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物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、未処理生物学的サンプル中のレベルと比較する。両サンプルの転写物レベルの差は、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。

10

【0230】

別の実施態様は、本発明のポリペプチド配列群を用いて或る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、個々に更なる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すなわちプロファイルは、所与の条件下で所与の時間に発現したタンパク質の数および相対存在量を定量することにより分析し得る。したがって、或る細胞のプロテオームのプロファイルは、特定の組織または細胞タイプのポリペプチドを分離および分析することにより作成し得る。或る実施態様では、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、次に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、通常はクーマシーブルー、あるいは銀染色液または蛍光染色液などの物質を用いてゲルを染色することにより、分散した、独自の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未処理のいずれかの生物学的サンプルから得られる同等に位置するタンパク質スポットの光学密度を比較し、処理に

20

30

【0231】

プロテオームのプロファイルは、IGSFPに特異的な抗体を用いてIGSFP発現レベルを定量することによっても作成可能である。或る実施態様では、マイクロアレイ上のエレメントとしてこれら抗体を用い、マイクロアレイをサンプルに曝して各アレイエレメントへのタンパク質結合レベルを検出することにより、タンパク質発現レベルを定量する（Lueking, A. 他 (1999) Anal. Biochem. 270:103-111、Mendoze, L.G. 他 (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と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。

40

【0232】

プロテオームレベルでの毒性シグネチャも中毒学的スクリーニングに有用であり、転写レベルでの毒性シグネチャと並行して分析するべきである。数種の組織の数種のタンパク質に対しては、転写物とタンパク質との存在量の相関が乏しいので（Anderson, N.L.及びJ. Seilhamer (1997) Electrophoresis 18:533-537）、転写物イメージには有意に影響しな

50

いがプロテオームのプロフィールを変化させるような化合物の分析においてプロテオーム毒物シグネチャは有用たり得る。更に、体液中の転写物の分析はmRNAの急速な分解のために困難なので、プロテオームのプロファイル作成はこのような場合により信頼し得、情報価値があり得る。

【0233】

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。処理された生体サンプル中で発現したタンパク質は、各タンパク質の量を定量し得るように分離する。各タンパク質の量を、非処理生物学的サンプル中の対応するタンパク質の量と比較する。両サンプルのタンパク質量の差は、処理サンプル中の試験化合物に対する毒性反応を示す。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。

10

【0234】

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。生体サンプルから得たタンパク質は、本発明のポリペプチドに特異的な抗体を用いてインキュベートする。抗体により認識されたタンパク質の量を定量する。処理された生物学的サンプル中のタンパク質の量を、非処理生物学的サンプル中のタンパク質の量と比較する。両サンプルのタンパク質量の差は、処理サンプル中の試験化合物に対する毒性反応を示す。

【0235】

マイクロアレイは、本技術分野で既知の方法を用いて調製し、使用し、そして分析する (Brennan, T.M. 他 (1995) の米国特許第5,474,796号、Schena, M. 他 (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619、Baldeschweiler 他 (1995) PCT出願第W095/251116号、Shalon, D. 他 (1995) PCT出願第W095/35505号、Heller, R.A. 他 (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155、Heller, M.J. 他 (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

【0236】

本発明の別の実施態様では、IGSFPをコードする核酸配列群を用いて、天然ゲノム配列をマッピングするのに有用なハイブリダイゼーションプローブ群を作製できる。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、コード配列よりも非コード配列が好ましい。例えば、多重遺伝子族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. 7:127-134、Trask, B.J. (1991) Trends Genet. 7:149--154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或る病状の遺伝を特定の染色体領域の遺伝または制限酵素断片長多型 (RFLP) と関連させるような遺伝子連鎖地図を発生させ得る。(例えば、Lander, E.S及びD. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357を参照)。

30

40

【0237】

蛍光原位置ハイブリッド形成法 (FISH) は、他の物理的及び遺伝地図データと関連し得る (例えば前出Meyersの965-968ページのHeinz-Ulrich, 他 (1995)を参照)。遺伝地図データの例は、種々の科学雑誌あるいはOnline Mendelian Inheritance in Man (OMIM) のウェブサイトに見ることができる。物理的な染色体地図上のIGSFPをコードする遺伝子の位置と特定の疾患との相関性、或いは特定の疾患に対する素因が、このような疾患と関連するDNA領域の決定に役立つため、更なる位置を決定するクローニングが行われる。

【0238】

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、正確な染色体上の遺伝子座が未知でも、関連するマーカー類をしばしば明らかにし得る。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。疾患または症候群に關与する遺伝子が、血管拡張性失調症の11q22-23領域のように、特定のゲノム領域への遺伝的連関によって大まかに位置決めがなされると、該領域にマップされた任意の配列は、更なる調査のための関連遺伝子或いは調節遺伝子を提示している可能性がある（例えばGatti, R.A. 他 (1988) Nature 336:577580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

【0239】

本発明の別の実施例では、IGSFP、その触媒作用断片或いは免疫原断片またはそのオリゴペプチドを、種々の任意の薬剤スクリーニング技術における化合物のライブラリのスクリーニングに用いることができる。薬剤スクリーニングに用いる断片は、溶液中に遊離しているか、固体支持物に固定されるか、細胞表面上に保持されるか、細胞内に位置しうる。IGSFPと検査する薬剤との結合による複合体の形成を測定してもよい。

【0240】

別の薬物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(Geysen, 他 (1984) PCT 出願番号 W084/03564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板上で合成する。試験用化合物は、IGSFP、或いはその断片と反応してから洗浄される。次に、本技術分野でよく知られている方法で、結合したIGSFPを検出する。精製したIGSFP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

【0241】

別の実施例では、IGSFPと特異結合可能な中和抗体がIGSFPとの結合について試験化合物と競合する、競合的薬剤スクリーニングアッセイを用いることができる。この方法では、抗体が、IGSFPと1つ以上の抗原決定因子を共有するどのペプチドの存在も検出する。

30

別の実施例では、IGSFPをコードするヌクレオチド配列を、将来に開発される分子生物学技術であり、現在知られているヌクレオチド配列の特性（限定はされないが、トリプレット遺伝コード、特異的な塩基対相互作用等を含む）に依存する新技術に用い得る。

【0242】

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。したがって、これ以下に記載する実施例は単なる例示目的にすぎず、いかようにも本発明を限定するものではない。

【0243】

前述した、および以下に記載する全ての特許出願、特許、刊行物、米国特許出願第60/275,249号、同第60/316,810号、同第60/323,977号、同第60/348,447および同第60/343,880号も含めて、言及することをもって本明細書の一部とする。

40

【実施例】

【0244】

1 cDNAライブラリの作製

Incyte cDNA群の由来は、LIFESEQ GOLDデータベース (Incyte Genomics, Palo Alto CA) に記載されたcDNAライブラリ群である。幾つかの組織はホモジナイズしてグアニジニウムイソチオシアネート溶液に溶解し、他の組織はホモジナイズしてフェノールにまたは変性剤群の好適な混合液に溶解した。混合液の1例であるTRIZOL (Life Technologies) は、フェノールとグアニジンイソチオシアネートとの単相溶液である。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

50

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。

【0245】

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, Chatsworth 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。

【0246】

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既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, 5.1-6.6ユニット等を参照)。逆転写は、オリゴd(T)またはランダムプライマーを用いて開始した。合成オリゴヌクレオチドアダプターを二本鎖cDNAに連結反応させ、好適な制限酵素または酵素でcDNAを消化した。殆どのライブラリに対して、cDNAのサイズ (300~1000bp) 選択は、SEPHACRYL S1000、SEPHAROSE CL2BまたはSEPHAROSE CL4Bカラムクロマトグラフィー (Amersham Pharmacia Biotech)、或いは分取アガロースゲル電気泳動法を用いて行った。cDNAは好適なプラスミドのポリリンカーの適合する制限酵素部位へ連結された。好適なプラスミドは、例えばPBLUESCRIPTプラスミド (Stratagene)、PSPORT1プラスミド (Life Technologies)、PCDNA2.1 プラスミド (Invitrogen, Carlsbad CA)、PBK-CMV プラスミド (Stratagene)、PCR2-TOPOTAプラスミド (Invitrogen)、PCMV-ICISプラスミド (Stratagene)、pIGEN (Incyte Genomics, Palo Alto CA)、pRARE (Incyte Genomics) またはpINCY (Incyte Genomics) 等およびその誘導体である。組換えプラスミドは、Stratagene社のXL1-Blue、XL1-BlueMRFまたはSOLR、或いはLife Technologies社のDH5、DH10BまたはELECTROMAX DH10Bを含む大腸菌細胞に形質転換した。

【0247】

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。プラスミドの精製には、下記の少なくとも1つを用いた。すなわち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およびQIAWELL 8 Ultra Plasmid 精製システム、R.E.A.L. Prep 96プラスミド精製キットのいずれかである。プラスミドは、沈殿させた後、0.1mlの蒸留水に再懸濁して、凍結乾燥して或いは凍結乾燥しないで4℃で保管した。

【0248】

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

【0249】

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

10

20

30

40

50

はMICROLAB 2200 (Hamilton) 液体転移システムと併用して処理した。cDNAのシーケンシング反応は、Amersham Pharmacia Biotech社が提供する試薬、またはABIシーケンシングキット、例えばABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems) の試薬を用いて準備した。cDNAのシーケンシング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シーケンシングシステム (Molecular Dynamics) が、標準ABIプロトコル及び塩基呼び出し (base calling) ソフトウェアを用いるABI PRISM 373または377シーケンシングシステム (Applied Biosystems) が、或いはその他の本技術分野で既知の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, 7.7ユニットに概説) を用いて決定した。cDNA配列の幾つかを選択して、実施例 8 に記載した方法で配列を伸長させた。

10

【0250】

Incyteの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、IncyteのcDNA配列、またはその翻訳を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) と、ヒト、ラット、マウス、線虫、出芽酵母 (*Saccharomyces cerevisiae*)、分裂酵母 (*Schizosaccharomyces pombe*)、および驚口瘡カンジダ (*Candida albicans*) からの配列を含むPROTEOMEデータベース (Incyte Genomics, Palo Alto CA)、及びPFAM等隠れマルコフモデル (HMM) に基づいたタンパク質ファミリーデータベース並びに、SMART (Schultz 他 (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864, Letunic, I. 他 (2002) Nucleic Acids Res. 30:242-244) のようなHMMに基づいたタンパク質ドメインデータベースから選択したデータベースに対して問い合わせた。(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) Cuff. Opin. Struct. Biol. 6:361-365等を参照)。問合せは、BLAST、FASTA、BLIMPS及びHMMERに基づくプログラムを用いて行った。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された。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、PROTEOMEデータベース、BLOCKS、PRINTS、DOMO、PRODOM及びProsites等のデータベース、PFAM、INCY、TIGRFAM等の隠れマルコフモデル (HMM) に基づいたタンパク質ファミリーデータベース、およびSMARTのようなHMMに基づいた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) を用いて分析される。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マルチシーケンシング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。

20

30

40

【0251】

Incyte cDNA及び完全長配列の分析及び構築に利用したツール、プログラム及びアルゴリズムの概略と、適用可能な説明、参照文献、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は適切な参照文献であり、全ての文献は全体を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、

50

確率値その他のパラメータを示す（スコアが高いほど、また確率値が低いほど、2配列間の同一性が高くなる）。

【0252】

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:13-24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2に示した。

【0253】

4 ゲノムDNAからのコード配列の同定及び編集

推定上の免疫グロブリンスーパーファミリータンパク質は、公共のゲノム配列データベース（例えば、gbpriやgbhtg）においてGenscan遺伝子同定プログラムを実行して初めに同定された。Genscanは、様々な生物からゲノムDNA配列を分析する汎用遺伝子同定プログラムである（Burge, C及びS. Karlin (1997) J. Mol. Biol. 268:78-94、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。これらのGenscan推定cDNA配列の内、どの配列が免疫グロブリンスーパーファミリータンパク質をコードするかを決定するために、コードされたポリペプチドをPFAMモデルにおいて免疫グロブリンスーパーファミリータンパク質について問合せて分析した。潜在的な免疫グロブリンスーパーファミリータンパク質はまた、免疫グロブリンスーパーファミリータンパク質として注釈が付けられていたIncyte cDNA配列への相同性を基に同定された。こうして選択されたGenscan予測配列は、次にBLAST分析により公共データベースgenpept及びgbpriと比較した。必要であれば、genpeptからのトップのBLASTヒット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、Genscanにより予測された配列のエラーを修正した。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、したがって転写の証拠を提供した。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、編集した、または非編集のGenscan予測コード配列に完全に由来する。

【0254】

5 cDNA配列データとのゲノム配列データの統合(assembly)

ステッチ配列(Stitched Sequence)

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

。Genscanにより予測された不正確なエキソンは、genpeptからのトップBLASTヒットと比較することにより修正した。必要な場合には、追加cDNA配列を用いるかゲノムDNAの検査により配列を更に伸長させた。

【0255】

ストレッチ配列 (Stretched Sequence)

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

10

【0256】

6 IGSFPをコードするポリヌクレオチドの染色体マッピング

SEQ ID NO:13-24を構築するために用いた配列を、BLAST及びSmith-Watermanアルゴリズムを用いて、Incyte LIFESEQデータベース及び公共のドメインデータベースの配列と比較した。SEQ ID NO:13-24と一致するこれらのデータベースの配列を、Phrapなどの構築アルゴリズム (表7) を使用して、連続及びオーバーラップした配列のクラスターに組み入れた。スタンフォード・ヒトゲノムセンター (SHGC)、ホワイトヘッド・ゲノム研究所 (WIGR)、Genethon等の公的な情報源から入手可能な放射線ハイブリッド及び遺伝地図データを用いて、クラスタ化された配列が既にマッピングされていたかを決定した。マッピングされた配列が或るクラスターに含まれている場合、そのクラスターの全配列が、個々の配列番号と共に、地図上の位置に割り当てられた。

20

【0257】

地図上の位置は、ヒト染色体の範囲または区間として表される。センチモルガン間隔の地図上の位置は、染色体のpアームの末端に関連して測定する。(センチモルガン (cM) は、染色体マーカー間の組換え頻度に基づく計測単位である。平均して、1 cMは、ヒト中のDNAの1メガベース (Mb) にほぼ等しい。尤も、この値は、組換えのホットスポット及びコールドスポットに起因して広範囲に変化する。) cM距離は、配列が各クラスターに含まれるような放射線ハイブリッドマーカーに対する境界となるようなGenethonによってマッピングされた遺伝マーカーに基づく。NCBI「GeneMap'99」 (<http://www.ncbi.nlm.nih.gov/genemap/>) などの一般個人が入手可能なヒト遺伝子マップおよびその他の情報源を用いて、既に同定されている疾患遺伝子群が、上記した区間内若しくは近傍に位置するかを決定できる。

30

【0258】

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. 他, 4章及び16章等を参照)。

40

【0259】

BLASTを適用する類似のコンピュータ技術を用いて、GenBankやLIFESEQ (Incyte Genomics) 等のcDNAデータベースにおいて同一または関連分子を検索した。ノーザン分析は、多数の膜系ハイブリダイゼーションよりも非常に速い。更に、特定の一致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

50

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。

【0260】

【数1】

(BLASTスコア×配列一致率)

5 × (長さ(配列1), 長さ(配列2))の最小値

【0261】

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のノーマライズされた値であり、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、各不一致塩基対に-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スコアのセグメント対を用いて積スコアを計算する。積スコアは、断片的オーバーラップと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%オーバーラップしているか、他端が88%一致し、100%オーバーラップ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%オーバーラップしているか、他端が79%一致し、100%オーバーラップしているかのいずれかの場合に得られる。

10

20

【0262】

或いは、IGSFPをコードするポリヌクレオチド配列は、由来する組織に対して分析する。例えば或る完全長配列は、Incyte cDNA配列(実施例3を参照)と少なくとも一部はオーバーラップするように構築される。各cDNA配列は、ヒト組織から作製されたcDNAライブラリに由来する。各ヒト組織は、以下の臓器/組織カテゴリーの1つに分類される。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/混合性または尿路である。各カテゴリーのライブラリ数を数えて、全カテゴリーの総ライブラリ数で除する。同様に、各ヒト組織は、以下の疾患/条件カテゴリー即ち癌、細胞株、発達、炎症、神経性、外傷、心血管、プール、その他の1つに分類される。各カテゴリーのライブラリ数を数えて、全カテゴリーの総ライブラリ数で除する。得られるパーセンテージは、IGSFPをコードするcDNAの組織特異的および疾患特異的な発現を反映する。cDNA配列およびcDNAライブラリ/組織の情報は、LIFESEQ GOLD データベース(Incyte Genomics, Palo Alto CA)から得ることができる。

30

【0263】

8 IGSFPをコードするポリヌクレオチドの伸長

完全長のポリヌクレオチド配列はまた、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。開始プライマーは、長さが約22~30ヌクレオチド、GC含有率が約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。

40

【0264】

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。

【0265】

50

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ウェルプレート内で行った。反応混合液は、鋳型DNA及び200nmolの各プライマー、 Mg^{2+} と $(NH_4)_2SO_4$ と2-メルカプトエタノールを含む反応バッファー、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。プライマー対、PCI AとPCI Bに対して以下のパラメータで増幅を行った。ステップ1: 94 で3分間、ステップ2: 94 で15秒間、ステップ3: 60 で1分間、ステップ4: 68 で2分間、ステップ5: ステップ2、3および4を20回繰り返す、ステップ6: 68 で5分間、ステップ7: 4 で保存。別法では、プライマー対であるT7とSK+とに対して以下のパラメータで増幅を行った。ステップ1: 94 で3分間、ステップ2: 94 で15秒間、ステップ3: 57 で1分間、ステップ4: 68 で2分間、ステップ5: ステップ2、3および4を20回繰り返す、ステップ6: 68 で5分間、ステップ7: 4 で保存。

10

【0266】

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 (Corning 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。

20

【0267】

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コレ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、pUC 18ベクター (Amersham Pharmacia Biotech) への再連結反応前に音波処理またはせん断した。ショットガン・シーケンシングのために、消化したヌクレオチドを低濃度 (0.6 ~ 0.8%) のアガロースゲル上で分離し、断片を切除し、寒天を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オーバーハングを満たし、大腸菌細胞に形質移入した。形質転換した細胞を選択して抗生物質を含む培地に移し、それぞれのコロニーを切りとってLB/2Xカルベニシリン培養液の384ウェルプレートに37 で一晩培養した。

30

【0268】

細胞を溶解して、Taq DNAポリメラーゼ (Amersham Pharmacia Biotech) 及びPfu DNAポリメラーゼ (Stratagene) を用いて以下の手順でDNAをPCR増幅した。ステップ1: 94 で3分間、ステップ2: 94 で15秒間、ステップ3: 60 で1分間、ステップ4: 72 で2分間、ステップ5: ステップ2、3および4を29回繰り返す、ステップ6: 72 で5分間、ステップ7: 4 で保存。上記したようにPICOGREEN試薬 (Molecular Probes) でDNAを定量化した。DNAの回収率が低いサンプルは、上記と同一の条件を用いて再増幅した。サンプルは20%ジメチルスルホキシド (1:2, v/v) で希釈し、DYENAMIC エネルギートランスファー シーケ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 シーケンシングレディ反応キット (Terminator cycle sequencing ready reaction kit) (Applied Biosystems) を用いてシーケンシングした。

40

【0269】

同様に、上記手順を用いて完全長ポリ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。

【0270】

9 IGSFPをコードするポリヌクレオチドの一塩基多型の同定
一塩基多型性 (SNP) として知られる一般的なDNA配列変異体は、LIFESEQデータベース

50

(Incyte Genomics)を用いてSEQ ID NO:13-24において同定された。実施例3に記述されているように同じ遺伝子からの配列は共にクラスター化され、構築され、遺伝子内の全ての配列変異体を同定することができた。一連のフィルタから成るアルゴリズムは、SNPを他の配列変異体から区別するために用いられる。前段フィルタ群が、最小限のPhredクオリティスコア15を要求することにより大多数のベースコールのエラーを除去し、また、配列アラインメントエラーと、ベクター配列、キメラ、スプライス変異体の、不適切なトリミングに起因するエラーを除去した。先進の染色体分析の自動化した手順により、推定上のSNPの近傍の本来のクロマトグラムファイルを分析した。クローンエラー-フィルタ群は、統計的に生み出されたアルゴリズムを用いて、逆転写酵素、ポリメラーゼあるいは体細胞性突然変異によって引き起こされる等の、実験プロセッシング中に導入されたエラーを同定した。クラスタリングエラー-フィルタ群は、統計的に生み出されたアルゴリズムを用いて、近縁の相同体あるいは偽遺伝子のクラスタリングに起因するエラー、または非ヒト配列によるコンタミネーションによるエラーを同定した。フィルタの最終セットは、免疫グロブリンまたはT細胞受容体に見出される重複とSNPを除去した。

10

【0271】

異なる4つのヒト集団のSNP部位における対立遺伝子頻度を分析するために、高処理MASARRAYシステム(Sequenom, Inc.)を用いる質量分析によって、更なる特徴付けのためにいくつかのSNPが選択された。白人集団は92人(男性46人、女性46人)から成り、そのうち83人はユタ州、4人はフランス、3人はベネズエラ、2人はアーミッシュの出身である。アフリカ系集団は194人(男性97人、女性97人)から成り、全てアフリカ系米国人である。ラテンアメリカ系集団は324人(男性162人、女性162人)から成り、全てメキシコ出身である。アジア系の集団は126人(男性64人、女性62人)からなり、中国人43%、日本人31%、コリアン13%、ベトナム人5%、他のアジア系8%の親の構成が報告されている。対立遺伝子頻度は最初に白人集団で分析された。この集団で対立遺伝子変異を示さないSNPの時には他の3つの集団で更に試験されない場合もあった。

20

【0272】

10 個々のハイブリダイゼーションプローブの標識化及び使用

SEQ ID NO:13-24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)等の最新技術のソフトウェアを用いて設計し、各オリゴマー50 pmolと、[^{-32}P]アデノシン3リン酸(Amersham Pharmacia Biotech) 250 μCi 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)を用いて実質的に精製する。下記のいずれか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 10^7 カウントの標識されたプローブを含むアリコットを用いる。すなわちAse I、Bgl II、Eco RI、Pst I、Xba I、またはPvu II (DuPont NEN)である。

30

【0273】

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

40

【0274】

11 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結合または合成は、フォトリソグラフィ、ピエゾ式印刷(インクジェット印刷、前出のBalteschweiler等を参照)、機械的マイクロ

50

スポットティング技術及びこれらから派生したものをを用いて達成することが可能である。上記各技術において基板は、均一な、無孔の表面を持つ固体とすべきである (Schna (1999) 前出)。推奨する基板には、シリコン、シリカ、スライドガラス、ガラスチップ及びシリコンウエハがある。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した手順を利用して、熱的、紫外線的、化学的または機械的結合手順を用いて基板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、利用可能な、当業者に公知の方法と機械とを用いて作製でき、任意の適正数のエレメントを有し得る (例えば Schna, M. 他 (1995) Science 270:467-470、Shalon, D. 他 (1996) Genome Res. 6:639-645、Marshall, A. および J. Hodgson (1998) Nat. Biotechnol. 16:27-31 を参照)。

【0275】

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が、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、LASERGENE 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の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でのハイブリダイゼーションを検出する。或いは、レーザー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製及び使用について、以下に詳述する。

【0276】

組織または細胞サンプルの調製

グアニジニウムチオシアネート法を用いて組織サンプルから全 RNA を単離し、オリゴ (dT) セルロース法を用いてポリ (A)⁺ RNA を精製する。各ポリ (A)⁺ RNA サンプルを、MMLV 逆転写酵素、0.05 pg/μl のオリゴ (dT) プライマー (21mer)、1 × 第一鎖合成バッファー、0.03 unit/μl の RNアーゼ阻害因子、500 μM の dATP、500 μM の dGTP、500 μM の dTTP、40 μM の dCTP、40 μM の dCTP-Cy3 (BDS) または dCTP-Cy5 (Amersham Pharmacia Biotech) を用いて逆転写する。逆転写反応は、GEMBRIGHT キット (Incyte) を用い、200 ng のポリ (A)⁺ RNA を含有する体積 25 ml で行う。特異的対照ポリ (A)⁺ RNA は、非コード酵母ゲノム DNA から *in vitro* 転写により合成する。37 °C で 2 時間インキュベートした後、各反応サンプル (1 つは Cy3、もう 1 つは Cy5 標識) は、2.5 ml の 0.5 M 水酸化ナトリウムで処理し、85 °C 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 SP IN 30 ゲル濾過スピナラム (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) を用いて精製する。結合後、2 つの反応サンプルは、1 ml のグリコーゲン (1 mg/ml) を用いて析出させたエタノール、60 ml の酢酸ナトリウム及び 300 ml の 100% エタノールである。サンプルは次に、SpeedVAC (Savant Instruments Inc., Holbrook NY) を用いて完全に乾燥させ、14 μl の 5 × SSC / 0.2% SDS 中で再懸濁する。

【0277】

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、クローン化した cDNA インサートを有するベクターを含有する細菌細胞から増幅する。PCR 増幅は、cDNA インサートの側面に位置するベクター配列に相補的なプライマーを用いる。30 サイクルの PCR で 1 ~ 2 ng の初期量から 5 μg 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。

【0278】

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、処理の間及び処理後に 0.1% の SDS 及びアセトン中で

10

20

30

40

50

超音波処理をかけ、蒸留水で十分に洗浄する。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、蒸留水中で十分に洗浄し、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のオーブンで硬化させる。

【0279】

米国特許第5,807,522号で説明されている方法を用いて、コーティングしたガラス基板にアレイエレメントを付加する。該特許は、引用を以って本明細書の一部となす。平均濃度100ng/ μ lのアレイエレメントDNA1 μ lを、高速ロボット装置 (robotic apparatus) により、開放型キャピラリープリンティングエレメント (open capillary printing element) に充填する。装置はここで、スライド毎に約5nlのアレイエレメントサンプルを加える。

10

【0280】

マイクロアレイには、STRATALINKER UV架橋剤 (Stratagene) を用いてUV架橋する。マイクロアレイは、室温において0.2% SDSで1度洗浄し、蒸留水で3度洗浄する。リン酸緩衝生理食塩水 (PBS) (Tropix, Inc., Bedford MA) 中の0.2% カゼイン中において60 °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% 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。

【0281】

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 SSC、0.2% 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 μ g含む9 μ 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1.8cm²のカバーガラスで覆う。アレイは、顕微鏡用スライドより僅かに大きい空洞を有する防水チェンバーに移す。チェンバーのコーナーに140 μ lの5 \times 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 SSC、0.1% 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 SSC) において各々45 °Cで10分間、3度洗浄して乾燥させる。

20

【0282】

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488nm、Cy5の励起のためには632nmでスペクトル線を発生し得るInnova 70混合ガス10 Wレーザ (Coherent, Inc., Santa Clara CA) を備えた顕微鏡で検出する。20 \times 顕微鏡対物レンズ (Nikon, Inc., 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ースキャンする。本実施例で用いる1.8cm \times 1.8cmのアレイは、20 μ mの解像度でスキャンする。

30

【0283】

2回の異なるスキャンで、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565nm、Cy5では650nmである。装置は両方の蛍光色素からのスペクトルを同時に記録し得るが、レーザ源において好適なフィルターを用いて、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。

40

【0284】

スキャンの感度は通常、既知濃度のサンプル混合体に添加されるcDNA対照種により生成されるシグナル強度を用いて校正する。アレイ上の特定の位置には相補的DNA配列が含まれ、その位置におけるシグナルの強度を、ハイブリダイズする種の重量比1:100,000に相関させる。異なる源泉 (例えば試験される細胞及び対照細胞など) からの2つのサンプルを

50

、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、較正するcDNAのサンプルを2つの蛍光色素で標識し、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。

【0285】

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

10

【0286】

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、GEMT00LS遺伝子発現分析プログラム(Incyte)である。

【0287】

例えば、SEQ ID NO:19はマイクロアレイ分析によって実証されるように、毒性研究で示差発現を示した。SEQ ID NO:19の発現は、未処理のC3A細胞に対して、様々な薬剤(ステロイド、ステロイドホルモン等)で処理したヒトC3A肝臓細胞株で少なくとも2倍減少した。ヒトC3A肝臓細胞株は、成長の強力な接触阻害に関して選択したHepG2/C3(肝臓腫瘍を患う15歳の男子から単離した肝臓癌細胞株)のクローン誘導体である。C3A細胞株は、成熟したヒト肝臓の*in vitro*モデルとして十分に確立されている(Mickelson 他(1995) *Hepatology* 22:866-875; Nagendra 他(1997) *Am J Physiol* 272:G408-G416)。肝臓代謝への影響は、薬剤の薬効を理解するために重要である。よって、SEQ ID NO:19は薬剤の薬効を理解するために有用である。

20

【0288】

1.2 相補的ポリヌクレオチド

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

30

【0289】

1.3 IGSFPの発現

IGSFPの発現および精製は、細菌若しくはウイルスを基にした発現系を用いて行うことができる。細菌でIGSFPを発現させるためには、抗生物質耐性遺伝子と、cDNA転写レベルを高める誘導性のプロモーターを含む好適なベクターにcDNAをサブクローニングする。このようなプロモーターの例には、lacオペレーター調節エレメントと併用する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)で誘発されるとIGSFPを発現する。真核細胞でのIGSFPの発現は、昆虫細胞株または哺乳動物細胞株に一般にバキュロウイルスとして知られている *Autographica californica*核多角体病ウイルス(AcMNPV)を感染させて行う。バキュロウイルスの非必須の

40

50

多角体遺伝子を、相同組換え或いは転移プラスミドの媒介を伴う細菌の媒介による遺伝子転移のどちらかによって、IGSFP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、強力なポリヘドリンプロモーターによって高いレベルのcDNAの転写が行われる。組換えバキュロウイルスは、多くの場合は *Spodoptera frugiperda* (Sf9) 昆虫細胞への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.他(1994) Proc. Natl. Acad. Sci. USA 91:3224-3227、Sandig, V.他(1996) Hum.Gene Ther. 7:1937-1945等を参照)。

【0290】

殆どの発現系では、IGSFPが、例えばグルタチオンSトランスフェラーゼ(GST)と、またはFLAGや6-Hisなどのペプチドエピトープ標識と合成された融合タンパク質となるため、未精製の細胞溶解物からの組換え融合タンパク質の親和性ベースの精製を素早く1回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)。精製の後、GST部分を特定の操作部位でIGSFPからタンパク質的に切断できる。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) を用いた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IGSFPを直接用いて以下の実施例16および17のアッセイを行うことができる。

【0291】

1.4 機能的アッセイ

IGSFP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IGSFPをコードする配列の発現によって評価する。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cDNAをサブクローニングする。選択されるベクターには、pCMVSPORTプラスミド (Life Technologies) 及びpCR 3.1プラスミド (Invitrogen, Carlsbad CA) が含まれ、どちらもサイトメガロウイルスプロモーターを有する。リポソーム製剤或いは電気穿孔法を用いて、5~10 μ gの組換えベクターをヒト細胞株、例えば内皮由来または造血由来の細胞株に一過的に形質移入する。更に、標識タンパク質をコードする配列を含む1~2 μ gのプラスミドを同時に形質移入する。標識タンパク質の発現により、形質移入細胞と非形質移入細胞を区別する手段が与えられる。また、標識タンパク質の発現によって、cDNAの組換えベクターからの発現を正確に予想できる。標識タンパク質は、例えば緑色蛍光タンパク質 (GFP; Clontech)、CD64またはCD64-GFP融合タンパク質から選択できる。自動化された、レーザー光学に基づく技術であるフローサイトメトリー (FCM) を用いて、GFPまたはCD64-GFPを発現する形質移入された細胞を同定し、その細胞のアポトーシス状態や他の細胞特性を評価する。FCMは、細胞死に先行するか或いは同時に発生する現象を診断する蛍光分子の取込を検出して計量する。このような現象として挙げられるのは、ヨウ化プロピジウムによるDNA染色によって計測される核DNA含量の変化、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粒度の変化、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、及びフルオレセイン抱合したアネキシン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。

【0292】

遺伝子発現に与えるIGSFPの影響は、IGSFPをコードする配列とCD64またはCD64-GFPのどちらかが形質移入された高度に精製された細胞集団を用いて評価することができる。CD64またはCD64-GFPは、形質移入された細胞表面で発現し、ヒト免疫グロブリンG (IgG) の保存領域と結合する。形質転換された細胞と形質転換されない細胞とは、ヒトIgGがCD64に対

する抗体のどちらかで被覆された磁気ビードを用いて分離することができる(DYNAL, Lake Success, NY)。mRNAは、当分野で周知の方法で細胞から精製することができる。IGSFp及び目的の他の遺伝子をコードするmRNAの発現は、ノーザン分析やマイクロアレイ技術で分析することができる。

【0293】

15 IGSFPに特異的な抗体の作製

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

【0294】

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

【0295】

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)。完全フロイントアジュバントにおいてオリゴペプチド-KLH複合体を用いてウサギを免疫化する。得られた抗血清の抗ペプチド活性及び抗IGSFp活性を検査するには、ペプチドまたはIGSFp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。

【0296】

16 特異的な抗体を用いる天然IGSFpの精製

天然IGSFp或いは組換えIGSFpを、IGSFpに特異的な抗体を用いるイムノアフィニティークロマトグラフィにより実質的に精製する。イムノアフィニティークラムは、CNBr-活性化S EPHAROSE(Amersham Pharmacia Biotech)のような活性化クロマトグラフィー用レジンと抗IGSFp抗体とを共有結合させることにより形成する。結合後に、製造者の使用説明書に従ってレジンをブロックし、洗浄する。

【0297】

IGSFpを含む培養液をイムノアフィニティークラムに通し、IGSFpを選択的に吸着できる条件で(例えば、界面活性剤の存在下において高イオン強度のバッファーで)そのカラムを洗浄する。そのカラムを、抗体とIGSFpとの結合を切るような条件で(例えば、pH 2~3のバッファー、或いは高濃度の尿素またはチオシアン酸塩イオンのようなカオトロピック塩で)溶出させ、IGSFpを回収する。

【0298】

17 IGSFPと相互作用する分子の同定

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

【0299】

或いは、IGSFp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。

10

20

30

40

50

【0300】

IGSF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大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判定できる(Nandabalan, K. 他(2000)米国特許第6,057,101号)。

【0301】

18 IGSFP活性の実証

IGSF活性の或るアッセイでは、IGSFが血清からの抗原類を認識し沈殿させる能力を測定する。この活性の測定は、定量沈降反応で成し得る(Golub, E. S. 他(1987) *Immunology: A Synthesis*, Sinauer Associates, Sunderland, MA, 113-115ページ)。IGSFを、当分野で既知の方法で同位体標識する。一定量の標識IGSFに種々の血清濃度を加える。IGSF-抗原複合体を溶液から沈殿させ、遠心分離で収集する。沈殿性IGSF-抗原複合体の量は、沈殿物中に検出される放射性同位元素の量に比例する。沈殿性IGSF-抗原複合体の量を、血清濃度に対してプロットする。異なった血清濃度の特徴的沈降曲線が得られ、沈殿性IGSF抗原複合物の量は最初は血清中濃度の増加に比例して増加し、等価点を頂点とし、その後は血清中濃度の増加に比例して減少する。このように、沈殿性IGSF-抗原複合体の量は、抗原の制限量と過剰量の両方に対する感受性によって特徴付けられるIGSF活性の測定量である。

【0302】

別法として、IGSF活性の或るアッセイは、細胞表面におけるIGSFの発現を測定する。IGSFをコードするcDNAを、非白血球細胞株に形質移入する。細胞表面タンパク質類をビオチンで標識する(de la Fuente, M.A. 他(1997) *Blood* 90:2398-2405)。IGSF特異抗体を用いて免疫沈降を行い、SDS-PAGEおよ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素と非標識免疫沈降素との比が、細胞表面に発現したIGSFの量に比例する。

【0303】

別法として、IGSF活性のアッセイは、IGSFの過剰発現によって誘発される細胞凝集の量を測定することで行われる。このアッセイにおいては、NIH3T3等の培養細胞にIGSFをコードするcDNAで形質移入する。このcDNAは強いプロモーターの制御下にある適切な哺乳動物発現ベクターに含まれている。緑色蛍光タンパク質(CLONTECH)などの蛍光標識タンパク質をコードするcDNAとの共形質移入を行うと、安定な形質移入体を同定するのに役立つ。形質移入された細胞と形質移入されない細胞で細胞の凝集(塊化)量を比較する。細胞の凝集量によりIGSFの活性を直接的に測定できる。

【0304】

当業者には、本発明の範囲及び精神から逸脱しない、本発明の記載した方法及びシステムの種々の修正および変更は自明であろう。本発明について説明するにあたり幾つかの実施例に関連して説明を行ったが、本発明の請求の範囲が、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修正は、下記の特許請求の範囲内にあるものとする。

【0305】

(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。

【0306】

表2は、本発明のポリペプチド群のGenBank識別番号と、最も近いGenBank相同体の注釈(annotation)と、PROTEOMEデータベース識別番号と、PROTEOMEデータベース相同体群の注釈とを示す。各ポリペプチドとその相同体が一致する確率スコアも併せて示す。

【0307】

表3は、予測されるモチーフ及びドメインを含む本発明のポリペプチド配列の構造的特徴

10

20

30

40

50

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。

【0308】

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。

【0309】

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。

【0310】

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。

【0311】

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

【0312】

【表1】

表1

Incyte プロジェクト ID	ポリペプチド SEQ ID NO:	Incyte ポリペプチド ID	ポリヌクレオチド SEQ ID NO:	Incyte ポリヌクレオチド ID	CA2 試薬
3855123	1	3855123CD1	13	3855123CB1	
4547188	2	4547188CD1	14	4547188CB1	90065916CA2
3939883	3	3939883CD1	15	3939883CB1	
3163819	4	3163819CD1	16	3163819CB1	3163819CA2
8518269	5	8518269CD1	17	8518269CB1	90110559CA2
1592646	6	1592646CD1	18	1592646CB1	
7500191	7	7500191CD1	19	7500191CB1	
7500099	8	7500099CD1	20	7500099CB1	2836421CA2
7682434	9	7682434CD1	21	7682434CB1	
2202389	10	2202389CD1	22	2202389CB1	
7503597	11	7503597CD1	23	7503597CB1	
7503603	12	7503603CD1	24	7503603CB1	

10

20

30

40

【 0 3 1 3 】
【 表 2 - 1 】

表2-1

ポリペプチド SEQ ID NO.	Incyte ポリペプチド ID	GenBank ID NO.またはプロテオーム ID NO.	確率スコア	注釈
1	3855123CD1	g14572521	2.00E-70	[ヒト] NEPH1 (Donoviel,D.B. 他. (2001) Mol. Cell. Biol. 21 (14), 4829-4836)
2	4547188CD1	g11071950	9.60E-121	[マウス] (AB048834) Fca/m 受容体(Shibuya,A. 他. (2001) Nat. Immunol. 1 (5), 441-446)
3	3939883CD1	g1136501	6.90E-35	[ラット] 表面タンパク質 MCA-32 (Pirozzi,G. 他. (1995) J. Immunol. 155 (12), 5811-5818)
4	3163819CD1	g9887089	6.50E-32	[マウス] リンパ球抗原 108 アイソフォーム I (Peck,S.R. 他. (2000) Immunogenetics 52 (1-2), 63-72)
4	3163819CD1	g15384841	1.00E-112	[ヒト] 活性化 NK 受容体(Bottino,C. 他. (2001) The Journal of experimental medicine. 194 (3), 235-246)
5	8518269CD1	g9887089	5.20E-62	[マウス] リンパ球抗原 108 アイソフォーム I (Peck,S.R. 他. (2000) Immunogenetics 52 (1-2), 63-72)
5	8518269CD1	g15384841	0.00E+00	[ヒト] 活性化 NK 受容体(Bottino,C. 他. (2001) The Journal of experimental medicine. 194 (3), 235-246)
6	1592646CD1	g18376829	1.00E-154	[ヒト] (AF391163) 破骨細胞関連受容体 hOscar-M2 (Kim,N. 他. (2002) J. Exp. Med. 195 (2), 201-209)
6	1592646CD1	g2645890	2.00E-32	[ヒト] [GSF1 (Mazzarella,R. 他. (1998) Genomics 48 (2), 157-162)
7	7500191CD1	g2078518	0	[ヒト] neogenin (Vielmetter,J. 他. (1997) Genomics 41 (3), 414-421)
8	7500099CD1	g10197717	7.40E-191	[ヒト] 細胞表面分子 Ly-9 (Tovar,V. 他. (2000) Immunogenetics 51 (10), 788-793)
9	7682434CD1	g586	1.20E-80	[ウツ] put. pre-OPCAM (AA 1 - 345) (Schofield,P.R. 他. (1989) EMBO J. 8 (2), 489-495)

【 0 3 1 4 】
【 表 2 - 2 】

表2-2

ポリペプチドSEQ ID NO:	Incyte ポリペプチド ID	GenBank ID NO:またはプロテオーム ID NO:	確率スコア	注釈
		336698 OPC ML	2.2E-81	[ヒト][受容体(シグナル伝達)[原形質膜] オピオイド結合細胞接着分子は、グリコシルホスファチジルイノシトール(GPI)アンカー神経細胞接着分子であり免疫グロブリンスーパーファミリーのメンバーであるラット Rn.11366 に対して類似性が高い (Struyk, A. F., 他. (1995) Cloning of neurotrimin defines a new subfamily of differentially expressed neural cell adhesion molecules. J. Neurosci 15:2141-2156, Lane, C. M. 他. (1992) Regulation of an opioid-binding protein in NG108-15 cells parallels regulation of delta-opioid receptors. Proc Natl Acad Sci U S A 89:11234-11238)。
		332056 Rn. 11366	5.3E-80	[ラット][受容体(シグナル伝達)[原形質膜] グリコシルホスファチジルイノシトール(GPI) アンカー神経細胞接着分子であり免疫グロブリンスーパーファミリーのメンバーである、オピオイド結合細胞接着分子 (Hachisuka, A., 他. (2000) Developmental expression of opioid-binding cell adhesion molecule (OBCAM) in rat brain Brain Res Dev Brain Res 122:183-191)。
		330088 Lsa mp	3.9E-77	[マウス][原形質膜] 神経細胞の選択的成長と軸索の標的化である役割を果たす Ig ファミリータンパク質のメンバーである辺縁系関連連膜タンパク質 (Pimenta, A. F., 他. (1996) cDNA cloning and structural analysis of the human limbic system-associated membrane protein (LAMP). Gene 170:189-195)。
11	7503597OD1	g14572521 598720 FLJ 10845	5.40E-158 6.6E-66	[ヒト] NEPH1 (Donoviel, D.B., 他. (2001) Mol. Cell. Biol. 21 (14), 4829-4836) [ヒト] 免疫グロブリン(Ig)ドメインを含むタンパク質は、グリコシルホスファチジルイノシトール(GPI)アンカー神経細胞接着分子であるオピオイド結合細胞接着分子であるラット Rn.11366 の領域に類似性の低い領域を有する。

10

20

30

40

表2-3

ポリペプチド SEQ ID NO:	Incyte ポリペプチド ID	GenBank ID NO:またはプロテオーム ID NO:	確率スコア	注釈
		340970 NPH S1	5.6E-21	[ヒト][「原形質膜、細胞接着」腎臓の糸球体で発現される免疫グロブリンファミリーのメンバーであるネフリンは、腎臓濾過障壁の機能または発達である役割を果たす可能性があり、対応する遺伝子の突然変異は先天性ネフロージ症候群を引き起こす (Ruotsalainen, V 他 (2000) Role of nephrin in cell junction formation in human nephrogenesis. Am. J. Pathol. 157:1905-1916)。

10

20

30

40

50

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	3855123CD1	442	S37 S51 S118 S129 S138 S171 S227 S236 S252 S366 S379 S385 S398 T48 T261 T306 T389 Y60	N162	シグナルペプチド:M198-C223	HMMER
2	4547188CD1	577	S39 S108 S189 S296 S301 S405 S482 S493 S525 T6 T38 T88 T234 T260 T271 T335 T349 T350 T437 T486 T524 T569 Y24	N212 N321	免疫グロブリンドメイン:G13-A64, G97-A165 膜貫通ドメイン:A193-A221 N-末端はサイトソル内 免疫グロブリン DM00001 Q08180 426-518:V80-D172 シグナルペプチド:M46-P63, M46-Q64, P33-P63	HMMER_PFAM TMAP BLAST_DOMO HMMER
					免疫グロブリンドメイン:G120-I200 膜貫通ドメイン:S39-P67 R495-R517 N-末端は細胞質内に ある。	HMMER_PFAM TMAP
					免疫グロブリン DM00001 P01833 41-120:H128-G201	BLAST_DOMO
					免疫グロブリン DM00001 P15083 41-120:H128-F208	BLAST_DOMO
					免疫グロブリン DM00001 P01832 28-125:G120-G201	BLAST_DOMO
					免疫グロブリン DM00001 S48841 41-120:H128-G201	BLAST_DOMO

【 0 3 1 7 】

【 表 3 - 2 】

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
3	3939883CD1	385	S4 S21 S99 S133 S214 S330 S373 T40 T60 T116 T162 T179 T181 T201 T226 T259 T296 T311 T342 Y236 Y355	N93 N102 N131 N193 N199 N224	signal_cleavage:M1-T38	SPSCAN
					シグナルペプチド:M1-G41	HMMER
					細胞内ドメイン:M1-K19, K293-F385	TMHMMER
					膜貫通ドメイン:F20-S39, L270-P292	TMHMMER
					細胞外ドメイン:T40-K269	TMHMMER
					免疫グロブリンドメイン:G91-A147, D182-A240	HMMER_PPFAM
					受容体 Fc 免疫グロブリン PD01270:T135-V171, R183-P211 P 値<1.3e-3	BLIMPS_PRODOM
					表面 タンパク質 MCA32 PD095298:L30-V164	BLAST_PRODOM
					血小板 内皮細胞 接着 前駆体 シグナル分子 PECAM1 CD31 抗原 PD150932:C68-P305	BLAST_PRODOM
					ロイシンジッパーモチーフ:L270-L291	MOTIFS
					細胞接着配列:R308-D310	MOTIFS
4	3163819CD1	221	S43 S52 S78 S143 S157 S180 T148 T188 T215 Y82	N26 N33 N50 N67 N92 N170 N192 N202	シグナルペプチド:M1-G15, M1-L19, M1-N21	HMMER
					細胞外ドメイン:M1-K114	TMHMMER
					膜貫通ドメイン:M115-L137	TMHMMER
					細胞内ドメイン:R138-V221	TMHMMER

【表 3 - 3】

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
5	8518269CD1	332	S106 S112 S116 S154 S163 S189 S254 S268 S291 T123 T259 T299 T326 Y107 Y193	N58 N87 N137 N144 N161 N178 N203 N281 N303 N313	signal_cleavage:M1-S21	SPSCAN
					シグナルペプチド:M1-G15, M1-Y19, M1-S21, M1-S23	HMMER
					細胞外ドメイン:M1-K225	TMHMMER
					膜貫通ドメイン:M226-L248	TMHMMER
					細胞内ドメイン:R249-V332	TMHMMER
					免疫グロブリンドメイン:G35-I111, T146-A197	HMMER_PFAM
					抗原 前駆体 シグナル 免疫グロブリンフォールド 糖タンパク T	BLAST_PRODUM
					細胞表面 CD2 膜貫通 PD010953:G32-S205	
6	1592646CD1	288	S122 S172 S232 S241 T75	N73 N170 N181	signal_cleavage:M1-T43	SPSCAN
					シグナルペプチド:M26-T43, M1-T43	HMMER
					免疫グロブリンドメイン:G168-Y222, G71-Y127	HMMER_PFAM
					受容体 NK 細胞 キラー 前駆体 シグナル 白血球 免疫グロブリン様 ナチュラル 抑制性 PD000659:H55-A193	BLAST_PRODUM
					アルファ 1B 糖タンパク 免疫グロブリンフォールド 糖タンパク質 血漿 PD138678:Y54-I240	BLAST_PRODUM

10

20

30

40

50

【 0 3 1 9 】

【 表 3 - 4 】

表3-4

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	7500191CD1	1450	S46 S64 S81 S156 S294 S451 S606 S620 S677 S731 S834 S939 S1087 S1137 S1203 S1281 S1283 S1291 S1327 S1328 S1385 S1407 S1423 T143 T212 T279 T311 T365 T371 T458 T532 T581 T603 T628 T759 T784 T808 T869 T873 T892 T924 T948 T1051	N73 N210 N326 N470 N489 N639 N715 N909 N1135 N1287	signal_cleavage:M1-A33	SPSCAN
			T1117 T1121 T1187 T1414 Y127 Y408 Y890		シグナルペプチド:M1-G30, M1-A33	HMMER
					フィブロネクチンタイプIIIドメイン: P539-T621, P633-T721, P954-S1044, P439-S525, P739-L821, P853-S942	HMMER_PFAM
					免疫グロブリンドメイン: G263-A322, G166-V223, G67-A131, S355-A412	HMMER_PFAM
					サイトソルドメイン:T1117-A1450 膜貫通ドメイン:L1094-C1116 非サイトソル内ドメイン:M1-M1093	TMHMMER
					受容体チロシンキナーゼクラスVタンパク質 BL00790:V450-F476, Y477-G520, S554-K579	BLIMPS_BLOCKS

10

20

30

40

50

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
					フィブロネクチンタイプ III リポトシグネチャ PR00014:T752-P761, A908-Y926, Y1028-P1042	BLIMPS_PRINTS
					腫瘍抑制因子 NEOGENIN タンパク質 DCC 前駆体 結腸直腸 糖タンパク質 免疫グロブリンファミリー PD041287:D1169-T1448 PD009999:C1116-P1172	BLAST_PRODOM
					NEOGENIN タンパク質 PD020198:M1-R66	BLAST_PRODOM
					腫瘍抑制因子 PD171136:E58-V133	BLAST_PRODOM
					免疫グロブリン DM00001 P43146 328-410:P341-Q420 DM00001 P43146 42-127:F55-I140	BLAST_DOMO
					フィブロネクチンタイプ III 反復 DM00007 P43146 935-1014:A945-D1025 DM00007 P43146 834-912:T846-N923	BLAST_DOMO
					TonB-依存性受容体タンパク質シグネチャ 1: M1-R5	MOTIFS
8	7500099CD1	551	S6 S17 S46 S128 S163 S179 S229 S316 S321 S400 S431 S453 S512 S524 T73 T122 T141 T142 T160 T192 T212 T252 T277 T438 T439 T487 Y335	N68 N95 N120 N169 N173 N285 N436	signal_cleavage:M1-G47	SPSCAN
					免疫グロブリンドメイン:S171-A224, G60-I133	HMMER_PPFAM

【 0 3 2 1 】

【 表 3 - 6 】

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
					サイトゾルドメイン:K387-T551 膜貫通ドメイン:L365-W386 非サイトゾル内ドメイン:M1-K364	TMHMMER
					抗原 LY9 前駆体 シグナル 膜貫通 糖タンパク質 免疫グロブリンファミリー PD126134:P359-P545	BLAST_PRODOM
					抗原 前駆体 シグナル 免疫グロブリンファミリー 糖タンパク質 細胞表面 OD2 膜貫通 PD010953:Y55-T243	BLAST_PRODOM
					免疫グロブリン DM0000 Q01965 139-210:M161-S232	BLAST_DOMO
					B-細胞表面 糖タンパク質 BLAST-1 DM03635 P10252 1-239:V31-S232 DM03635 P1818 1-239:L32-S232	BLAST_DOMO
9	7682434CD1	336	S37 S175 S203 S207 S225 S282 S303 T43 T91 T143 T165 T219 T269 T290	N41 N49 N67 N137 N280 N288	signal_cleavage:M1-S30	SPSCAN
					シグナルペプチド:M1-R26, M1-S30	HMMER
					免疫グロブリンドメイン:G231-A293, G47-F114, G147-T197	HMMER_PFAM
					前駆体 シグナル 糖タンパク質 免疫グロブリンファミリー 細胞接着 GPI アンカー タンパク質 代替 PD005605:F35-Q124	BLAST_PRODOM

10

20

30

40

50

【表 3 - 7】

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	分析方法及びデータベース
10	2202389CD1	241	S44 S88 S112 S163 T10 T134 Y39	N75 N94 N110 N213	免疫グロブリン DM00001 P32736 39-125:D40-T123 DM00001 P32736 139-212:V136-D206 DM00001 P32736 226-306:I220-A302 signal_cleavage:M1-I25	BLAST_DOMO SPSCAN
11	7503597CD1	766	S159 S207 S215 S272 S373 S387 S454 S465 S474 S507 S551 S560 S576 S690 S703 S709 S722 T230 T301 T384 T585 T630 T713 Y48 Y307 Y396	N167 N253 N324 N498	免疫グロブリンドメイン:G51-A117 シグナルペプチド:M1-E19, M1-G21, M1-Q23, M1-L22	HMMER_PFAM HMMER
					免疫グロブリンドメイン:G62-A129, G163-A229, G433-A501, D264-V316, G349-A400 サイトソルドメイン:C547-V766 膜貫通ドメイン:V524-F546 非サイトソル内ドメイン:M1-A523 糖タンパク質 抗原 前駆体 PD02327:L141-I152, T169-I190	HMMER_PFAM TMHMMER
					不規則 CHIASM CROUGHEST タンパク質 前駆体 IRREC 膜貫通 免疫グロブリンフォールド 糖タンパク質 シグナル細胞 接着 PD124347:F50-V256, V261-E315	BLIMPS_PRODOM BLAST_PRODOM

10

20

30

40

50

【 0 3 2 3 】

【 表 3 - 8 】

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
12	7503603CD1	T6 Y24			免疫グロブリン DM0001 Q08180 31-126:S51-I142 ロイシンジッパー:L8-L29	BLAST_DOMO MOTIFS

10

20

30

40

50

【表 4 - 1】

表 4-1

ポリヌクレオチド SEQ ID NO/ Incyte ポリヌクレオチド ID/ 配列長	配列断片
13/3855123CB1/ 2691	1-751, 214-969, 282-871, 354-864, 370-934, 388-852, 390-848, 417-958, 459-874, 496-1309, 560-1313, 575-1092, 609-1298, 662-1207, 662-1300, 671-1317, 729-1212, 731-1313, 805-1157, 884-1300, 915-1566, 916-1566, 1010-1503, 1018-1306, 1038-1273, 1038-1641, 1070-1566, 1143-1524, 1231-1791, 1280-1566, 1499-2164, 1559-2094, 1559-2130, 1559-2199, 1559-2220, 1744-1982, 1752-2366, 1770-2621, 1785-2365, 1785-2576, 1815-2098, 1815-2355, 1818-2268, 1826-2440, 1846-2604, 1892-2548, 1898-2664, 1958-2234, 1958-2453, 1969-2691, 1976-2621, 1977-2691, 1985-2631, 1992-2564, 1993-2691, 2014-2625, 2017-2625, 2039-2625, 2081-2689,
14/4547188CB1/ 2518	2104-2665, 2109-2444, 2140-2691, 2175-2561, 2200-2691, 2217-2691, 2233-2677, 2309-2691, 2338-2691, 2360-2675, 2444-2691, 2505-2691, 1-148, 1-606, 1-762, 40-275, 147-553, 533-897, 736-1036, 736-1069, 861-1483, 870-1149, 879-1056, 900-1427, 920-1229, 923-1356, 1020-1643, 1049-1251, 1049-1312, 1049-1654, 1110-1537, 1114-1678, 1145-1500, 1152-1687, 1156-1687, 1190-1458, 1190-1836, 1199-1828, 1200-1793, 1222-1768, 1251-1817, 1300-1453, 1321-1944, 1321-1966, 1344-1925, 1360-2011, 1364-2011, 1397-1864, 1420-1930, 1422-2081, 1429-2001, 1438-1864, 1470-1522, 1487-1522, 1496-2172, 1500-1552, 1516-1697, 1544-2056, 1596-2090, 1597-2204, 1599-1803, 1602-2252, 1603-2197, 1642-2219, 1669-2081, 1686-2215, 1722-2336, 1767-2432, 1786-2448, 1813-2313, 1827-2479, 1833-2266, 1850-2021, 1863-2465, 2015-2210, 2015-2497, 2015-2499, 2034-2315, 2034-2507, 2034-2518, 2045-2430, 2162-2378
15/3939883CB1/ 1522	1-274, 1-467, 71-510, 124-727, 124-753, 124-794, 124-827, 124-892, 140-599, 180-982, 264-821, 274-929, 496-799, 496-954, 517-1052, 602-1075, 613-727, 717-1177, 796-887, 801-1307, 801-1485, 975-1231, 992-1522, 995-1480, 1059-1498, 1073-1522
16/3163819CB1/ 1084	1-287, 1-496, 1-502, 1-646, 1-650, 1-660, 1-970, 67-982, 97-694, 187-1084, 470-686
17/8518269CB1/ 1463	1-511, 1-804, 17-844, 27-305, 27-511, 38-511, 43-373, 43-511, 63-759, 119-511, 146-511, 147-511, 446-1361, 476-1073, 566-1463

【 0 3 2 5 】

【表 4 - 2】

表4-2

ポリヌクレオチド SEQ ID NO/ Incyte ポリヌクレオチド ID/ 配列長	配列断片
18/1592646CB1/ 1557	1-758, 40-1554, 182-563, 183-738, 277-841, 360-876, 360-877, 494-694, 714-980, 853-1459, 863-1394, 886-1281, 900-1178, 909-1454, 937-1489, 949-1225, 960-1203.
19/7500191CB1/ 5553	960-1436, 964-1245, 964-1250, 990-1552, 992-1221, 992-1531, 995-1253, 996-1535, 1010-1522, 1025-1527, 1089-1381, 1091-1310, 1119-1345, 1119-1528, 1119-1546, 1138-1355, 1140-1410, 1140-1465, 1140-1553, 1150-1555, 1158-1391, 1249-1557 1-500, 30-501, 214-734, 216-616, 216-618, 216-733, 216-768, 216-815, 216-877, 216-901, 219-765, 226-593, 226-617, 226-698, 228-759, 278-587, 278-697, 278-698, 278-740, 282-728, 282-729, 308-857, 316-876, 317-434, 455-988, 611-1167, 611-1202, 656-1263, 671-992, 683-1210, 738-1018, 1058-1319, 1127-5470, 1171-1303, 1193-1700, 1214-1830, 1223-1830, 1255-1915, 1333-1818, 1343-1625, 1438-2117, 1439-1910, 1450-1914, 1465-1910, 1493-1630, 1509-1910, 1515-2242, 1552-2101, 1606-2156, 1670-1782, 1738-2363, 1780-2114, 1780-2313, 1814-2267, 1859-1978, 1886-2453, 1895-2489, 1910-2565, 1942-2600, 2049-2396, 2114-2693, 2243-2484, 2421-2762, 2453-2637, 2665-3303, 2722-3012, 2731-3272, 2735-2970, 2778-3352, 2798-3240, 2819-3125, 2910-3495, 2971-3570, 3001-3281, 3050-3629, 3112-3767, 3147-3428, 3147-3715, 3201-3446, 3201-3568, 3201-3603, 3201-3698, 3201-3763, 3205-3659, 3239-3473, 3280-3894, 3280-3895, 3289-3949, 3419-4085, 3464-3693, 3476-4043, 3491-3784, 3492-4014, 3506-3987, 3546-4147, 3611-4177, 3620-4185, 3628-3868, 3656-4218, 3679-3889, 3679-4107, 3681-4028, 3690-4093, 3714-4354, 3719-4240, 3726-4245, 3773-4022, 3784-4262, 3797-4056, 3798-4065, 3855-3979, 3872-4089, 3889-4393, 3930-4543, 3947-4186, 3998-4654, 4005-4497, 4007-4201, 4017-4613, 4033-4246, 4033-4563, 4053-4417, 4053-4441, 4059-4716, 4065-4531, 4066-4372, 4073-4530, 4081-4419, 4088-4496, 4089-4678, 4148-4688, 4150-4681, 4195-4420, 4195-4431, 4195-4478, 4195-4708, 4195-4758, 4195-4828, 4204-4515, 4219-4809, 4229-4468, 4245-4745, 4245-4888, 4252-4878, 4255-4507, 4256-4500, 4274-4491, 4280-5057, 4281-4833, 4306-4866, 4314-4846, 4330-4828, 4330-4899, 4336-4805, 4358-4635, 4425-4880, 4426-4709, 4429-4770, 4429-4774, 4432-4952, 4444-4665, 4445-5152, 4506-5025, 4512-5096, 4590-5188, 4599-5150, 4602-5114, 4618-5153, 4651-4800, 4651-5127, 4652-5151, 4722-5146, 4736-5016, 4741-5213,

【 0 3 2 6 】
【 表 4 - 3 】

表4-3

ポリスクレオチド SEQ ID NO/ Inocyte ポリスクレオチド ID/ 配列長	配列断片
	4746-4970, 4746-5207, 4746-5213, 4747-4965, 4750-5165, 4752-5213, 4754-4981, 4754-5215, 4756-5159, 4771-5217, 4779-5213, 4785-5076, 4787-5164, 4790-5032, 4811-5092, 4812-5211, 4818-5217, 4835-5100, 4844-5158, 4844-5165,
	4854-5164, 4854-5217, 4863-5006, 4863-5197, 4863-5212, 4867-5163, 4867-5165, 4870-5057, 4881-5089, 4883-5138, 4885-5168, 4888-5149, 4907-5270, 4965-5219, 4966-5154, 4966-5217, 4991-5219, 4993-5164, 5050-5213, 5111-5217, 5283-5553, 5285-5525
20/7500099CB1/ 1849	1-270, 1-375, 1-400, 1-475, 1-534, 1-536, 4-632, 4-1847, 22-280, 28-293, 51-642, 123-754, 307-902, 411-1001, 434-969, 437-902, 447-1073, 450-1063, 485-1003, 487-1082, 498-988, 557-1131, 571-1069, 580-1191, 642-801, 729-1012, 810-1094,
	810-1098, 810-1102, 879-1039, 879-1391, 1100-1420, 1100-1628, 1100-1639, 1100-1647, 1100-1671, 1100-1775, 1101-1673, 1170-1782, 1241-1653,
	1251-1849, 1264-1849, 1271-1654, 1274-1704, 1345-1732, 1464-1757, 1475-1807, 1483-1738, 1615-1821
21/7682434CB1/ 1427	1-575, 72-612, 260-466, 341-923, 341-944, 341-974, 341-976, 341-1072, 371-891, 373-934, 632-1085, 676-935, 676-1093, 759-1275, 856-1156, 856-1427
22/2202389CB1/ 1014	1-365, 1-510, 246-507, 246-739, 336-990, 509-1013, 550-1014, 556-1014, 574-1014, 593-1014, 599-1014, 605-1014, 734-1007, 769-899

【 0 3 2 7 】
【 表 4 - 4 】

10

20

30

40

表4-4

ポリヌクレオチド SEQ ID NO/ Incyte ポリヌクレオチド ID/ 配列長	配列断片
23/7503597CB1/ 3695	1-642, 1-803, 1-820, 26-820, 54-590, 71-625, 197-821, 528-820, 618-1141, 618-1162, 620-806, 620-820, 849-1210, 922-1166, 922-1210, 945-1723, 946-1385, 946-1524, 946-1621, 946-1643, 946-1649, 946-1687, 946-1692, 970-1466, 1001-1770, 1006-1781, 1012-1354, 1015-1433, 1046-1792, 1046-3681, 1057-1524, 1060-1526, 1060-1846, 1205-1972, 1208-1731, 1211-1845, 1440-2053, 1920-2563, 1921-2563, 2015-2508, 2043-2278, 2043-2537, 2043-2599, 2043-2646, 2043-2657, 2043-2703, 2043-2759, 2043-2792, 2043-2872, 2044-2721, 2075-2563, 2236-2796, 2285-2563, 2473-3226, 2521-3367, 2572-3099, 2572-3135, 2572-3204, 2572-3225, 2572-3232, 2621-3523, 2733-3345, 2749-2988, 2775-3626, 2823-3103, 2823-3360, 2824-3273, 2831-3445, 2853-3629, 2917-3607, 2920-3695, 2947-3695, 2951-3629, 2957-3291, 2964-3239, 2964-3458, 2967-3630, 2976-3504, 2981-3626, 2983-3692, 2984-3629, 2984-3636, 2995-3569, 2995-3693, 3010-3630, 3019-3630, 3029-3695, 3036-3425, 3044-3630, 3048-3521, 3049-3693, 3077-3695, 3109-3670, 3117-3695, 3126-3695, 3145-3676, 3180-3566, 3204-3695, 3205-3680, 3220-3695, 3222-3695, 3238-3681, 3255-3695, 3311-3695, 3314-3695, 3343-3695, 3365-3680, 3449-3695
24/7503603CB1/ 2403	1-212, 1-818, 1-829, 1-2397, 612-758, 612-815, 612-1152, 622-911, 737-1359, 746-1024, 755-930, 776-1303, 789-902, 797-1100, 897-1519, 925-1127, 925-1188, 925-1530, 990-1554, 1022-1376, 1066-1334, 1066-1712, 1075-1704, 1076-1669, 1127-1693, 1197-1820, 1197-1838, 1220-1800, 1236-1889, 1259-2127, 1264-2127, 1273-1740, 1296-1805, 1299-2127, 1314-1740, 1328-2127, 1346-1398, 1350-2127, 1376-1428, 1377-2049, 1420-1934, 1470-2324, 1472-1968, 1473-2081, 1475-1679, 1478-2130, 1479-2075, 1513-2126, 1514-2323, 1531-2326, 1545-1959, 1562-2093, 1579-2325, 1599-2214, 1600-2326, 1623-2127, 1643-2310, 1662-2326, 1689-2191, 1703-2357, 1709-2145, 1739-2343, 1845-2295, 1869-2317, 1895-2382, 1946-2273, 2040-2256, 2104-2403

10

20

30

40

50

表 5

ポリヌクレオチド SEQ ID NO:	Incyte プロジェクト ID	代表的ライブラリ
13	3855123CB1	BRAHNON05
14	4547188CB1	COLXTDT01
15	3939883CB1	SKINBIT01
16	3163819CB1	TLYMTXT04
17	8518269CB1	TLYJTXF01
18	1592646CB1	EOSIHET02
19	7500191CB1	BRAIFER05
20	7500099CB1	LUNGDIN02
21	7682434CB1	BRABDIK02
22	2202389CB1	SPLNFET02
23	7503597CB1	BRAHNON05
24	7503603CB1	COLXTDT01

【 0 3 2 9 】

【 表 6 - 1 】

10

20

30

40

表6-1

ライブラリ	バクター	ライブラリの説明
BRABDIK02	PSPORT1	この増幅し正規化したライブラリの作製には、3人の提供者からプールしたcDNAを用いた。cDNAは、肺炎のため死亡した79才の白人女性(提供者A)、鬱血性心不全のため死亡した83才の白人男性(提供者B)、および食道癌のため死亡した87才の白人女性(提供者C)から採取した罹患小脳虫部組織から単離されたmRNAを使って作製された。病理検査によると、重度のアルツハイマー病(提供者A、B)および中度のアルツハイマー病(提供者C)が示された。患者歴は、提供者Aにおいて緑内障、偽水晶体、胃腸出血を伴う胃炎、末梢血管病、慢性閉塞性肺疾患、てんかん発作、緩解期のタバコ乱用、および一過性虚血発作があり、提供者Bにおいては、パーキンソン病およびアテローム硬化が、提供者Cには高血圧、冠動脈疾患、脳血管事故、および甲状腺機能低下がある。家族歴には、アルツハイマー病(提供者Aの母および兄弟姉妹)がある。この増幅したライブラリからの独立クローン群は、Soares 他, PNAS (1994) 91:9228-9232 及び Bonaldo 他, Genome Research 6 (1996):791 を適用した条件を用いて1ラウンドで標準化した。ただし、有意に長い(1ラウンドが48時間)再アニーリングハイブリダイゼーションを用いた。
BRAHNON05	pINCY	この標準化した海馬組織ライブラリは、海馬組織ライブラリの160万個の独立クローンから作製した。開始RNAは、心不全で死亡した35才の白人男性から採取した後部海馬から作製した。病理検査では、中度の軟膜線維症と大脳新皮質の複数微小梗塞が見られた。大脳半球は、焦点石灰沈着を伴う、中度の軟膜線維症と判明した。大脳半球全体に、縮んだわずかに好酸性の錐体ニューロンの形跡があった。神経膠症を伴う、顕微鏡的小部分の空洞化が大脳皮質全体に散在的に見られた。患者の病歴には、心筋症、CHF、心肥大、脾臓と肝臓の肥大がある。患者の使用薬剤には、シメチコン、ラシックス、ジゴキシン、コラース、ザンタック、キャプトプリルおよびバゾテックが含まれる。このライブラリは、極めて長時間(48時間/1回)の再アニーリングによるハイブリダイゼーションを用いた他は、Soares 他, PNAS (1994) 91:9228 および Bonaldo 他, Genome Research 6(1996) 791 を適用した条件を用いて2回にわたり標準化した。
BRAIFER05	pINCY	ライブラリは妊娠23週目で死産の左心低形成症胎児(白人男子)から採取した脳組織から単離したRNAを用いて作製した。
COLXTDT01	pINCY	ライブラリは37才の黒人女性の筋腫摘出術、掻爬、右卵管采領域(fimbrial region)生検、付随的虫垂切除術時に取り除かれた盲腸の結腸組織から単離したRNAを用いて作製した。病理学的検査では通常の盲腸であった。随伴する腫瘍組織の病理学検査では、複数(12)の子宮平滑筋腫が見られた。患者の病歴は月経閉止前の月経過多および肺のサルコイドーシスを含んでいた。家族の病歴には、急性心筋梗塞およびアテローム硬化型冠動脈疾患が含まれる。
EOSIHET02	PBLUESCRIPT	ライブラリは48才の白人男性からアフェレーンされた末梢血液細胞から単離したRNAを用いて作製した。患者の病歴には、過好酸球増加症がある。ライト染色液によって細胞集団の77%以上が好酸球であることが判明した。
LUNGDI02	pINCY	この標準化した肺組織ライブラリは、罹患肺組織ライブラリの76万個の独立クローンから作製した。開始RNAは罹患肺組織から単離されたRNAを使って作製された。病理は特発性肺疾患を示した。このライブラリは、極めて長時間(48時間/1回)の再アニーリングによるハイブリダイゼーションを用いたことを除いては、Soares 他, PNAS (1994) 91:9228-9232 および Bonaldo

10

20

30

40

50

表6-2

ライブラリ	ベクター	ライブラリの説明
SKINBIT01	pINCY	他, Genome Research (1996) 6:791 を適応した条件を用いて 2 回にわたリノーマライズした。 ライブラリは左下腿の罹患皮膚組織から単離した RNA を用いて作製した。患者の病歴には左下腿の結節性紅斑がある。
SPLNFET02	pINCY	ライブラリは妊娠 23 週で死亡の白人男子胎児から採取した脾臓組織から単離した RNA を用いて作製された。
TLYJTXF01	PRARE	この 5 キャップの単離された完全長ライブラリは、ある男性の T 細胞由来の処理した Jurkat 細胞株から単離した RNA を用いて作製された。細胞は 5nM の PMA と 50ng/ml のイオノマイシンで 1 時間処理した。患者の病歴には急性 T 細胞白血病が含まれる。
TLYMTXT04	pINCY	ライブラリは供与者のプールから得た CD4+ T 細胞から単離された RNA を用いて作製した。この細胞は CD3 と CD28 抗体で処理した。

10

20

30

40

50

表7-1

プログラム	説明	参考文献	パラメータ閾値
ABI FACTURA	ベクター配列を除き、核酸配列内のまぎらわしい塩基をマスクするプログラム	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	1種のFast Data Finderであり、アミノ酸配列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。アミノ酸配列および核酸配列の配列類似性検索に有用であり、BLASTIには5つの機能がある:blastp、blastn、blastx、tblastnおよびtblastx。	Altschul, S.F. 他 (1990) J. Mol. Biol. 215:403-410, Altschul, S.F. 他 (1997) Nucleic Acids Res. 25:3389-3402.	EST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は最低5つの機能からなる:fasta、tfasta、fastx、tfastxおよびssearch。	Pearson, W.R. 及びD.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448, Pearson, W.R. (1990) Methods Enzymol. 183:63-98, Smith, T.F. 及びM.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	EST:fasta E 値=1.0E-6 構築された EST:fasta 同一性=95%以上 一致した長さ=200塩基以上 fasta E 値=1.0E-8以下 完全長配列: fastx スコア=100以上
BLIMPS	配列をBLOCKS、PRINTS、DOMO、PRODOMおよびPFAM データベースの配列と対応させて遺伝子ファミリー、配列相同性および構造的アンギャープリント領域を検索するBLocks IMProved サーチャー。	Henikoff, S. 及びJ.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572, Henikoff, J.G. 及び S. Henikoff (1996) Methods Enzymol. 266:88-105, 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センサ配列の隠れマルコフモデル(HMM)に基づいたデータベースに対して問合せ配列を検索するアルゴリズム。	Krogh, A. 他 (1994) J. Mol. Biol. 235:1501-1531, Sonnhammer, E.L.L. 他 (1988) Nucleic Acids Res. 26:320-322, Durbin, R. 他 (1998) Our World View, in a Nutshell, Cambridge Univ. Press, 1-350ページ。	PFAM ヒット:確率値=1.0E-3 以下 シングル ベンチド ヒット:スコア=0以上

表 7-2

プログラム	説明	参考文献	パラメータ閾値
ProfileScan	Prositeで定義された配列パターンと一致するタンパク質配列内の構造的モチーフおよび配列モチーフを検索するアルゴリズム。	Gribskov, M. 他 (1988) CABIOS 4:61-66, Gribskov, M. 他 (1989) Methods Enzymol.183:146-159, Bairoch, A. 他 (1997) Nucleic Acids Res. 25:217-221.	ノーマライズされた質のスコア \geq その特定のPrositeモチーフに対するGCG指定"HIGH"値 通常、スコア=1.4~2.1.
Phred	高い感度と確率で自動配列決定機のトレースを調べるベークーリングアルゴリズム。	Ewing, B. 他 (1998) Genome Res. 8:175-185, Ewing, B及びP. Green (1998) Genome Res. 8:186-194.	
Phrap	Smith-Waterman アルゴリズムの効率的実装に基づいたプログラムである、SWATとCrossMatchを含むPhils 改訂構築プログラムで、配列相溶性検索やDNA配列の構築に有用である。	Smith, T.F.およびM.S. Waterman (1981) Adv. Appl. Math. 2:482-489, Smith, T.F.およびM.S. Waterman (1981) J. Mol. Biol. 147:195-197, Green, P., University of Washington, Seattle, WA.	スコア=120以上 一致した長さ=56以上
Consed	Phrap で構築したものの表示、編集用グラフィカルツール。	Gordon, D. 他 (1998) Genome Res. 8:195-202.	
SPScan	タンパク質配列をスキャンして、分泌シグナルペプチドの存在を調べる重み行列解析プログラム。	Nielson, H. 他 (1997) Protein Engineering 10:1-6, Claverie, J.M.及びS. Audic (1997) CABIOS 12:431-439.	スコア=3.5以上
TMAP	重み行列を用いてタンパク質配列上の膜貫通セグメントを描写し、配向を決定するプログラム。	Persson, B.およびP.Argos (1994) J. Mol. Biol. 237:182-192, Persson, B.およびP. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	隠れMarkov モデル(HMM)を使ってタンパク質配列上の膜貫通セグメントを描写し、配向を決定するプログラム。	Sonnhammer, E.L. 他 (1998) Proc.Sixth Intl. Conf. on Intelligent Systems for Mol.Biol., Glasgow 他, 編集, The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, 175-182ページ。	
Motifs	Prositeで定義された配列と一致したパターンのアミノ酸配列を検索するプログラム。	Bairoch, A. 他 (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, 第9版, M51-59ページ, Genetics Computer Group, Madison, WI.	

【国際公開パンフレット】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
19 September 2002 (19.09.2002)

PCT

(10) International Publication Number
WO 02/072794 A2

- (51) International Patent Classification: C12N
- (21) International Application Number: PCT/US02/09052
- (22) International Filing Date: 12 March 2002 (12.03.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/275,249 12 March 2001 (12.03.2001) US
60/316,810 31 August 2001 (31.08.2001) US
60/523,977 21 September 2001 (21.09.2001) US
60/548,447 26 October 2001 (26.10.2001) US
60/543,880 2 November 2001 (02.11.2001) 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): YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). XU, Yuning [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). THANGAVELU, Kavitha [IN/US]; 1950 Montecito Avenue #23, Mountain View, CA 94043 (US). WARREN, Bridget, A. [US/US]; 10130 Parkwood Drive # 2, Cupertino, CA 95014 (US). TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). DUGGAN, Brendan, M. [AU/US]; 243 Buena Vista Avenue # 306, Sunnyvale, CA 94086 (US). TRAN, Uyen, K. [US/US]; 2638 Mabury Square, San Jose, CA 95133 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (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, EC, EE, ES, FI, GB, GD, GE, GL, GM, GR, GU, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LU, LV, MA, MD, MG, MK, MN, MW, MX, MY, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KI, LS, MW, MZ, SD, SL, SZ, UG, ZM, 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, NI, SN, TD, TG).
- Published:
without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 02/072794 A2

(54) Title: IMMUNOGLOBULIN SUPERFAMILY PROTEINS

(57) Abstract: The invention provides human immunoglobulin superfamily proteins (IGSF) and polynucleotides which identify and encode IGSF. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of IGSF.

WO 02/072794

PCT/US02/09052

IMMUNOGLOBULIN SUPERFAMILY PROTEINS**TECHNICAL FIELD**

This invention relates to nucleic acid and amino acid sequences of immunoglobulin superfamily
5 proteins and to the use of these sequences in the diagnosis, treatment, and prevention of immune
system, neurological, developmental, muscle, and cell proliferative disorders, and in the assessment of
the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of
immunoglobulin superfamily proteins.

BACKGROUND OF THE INVENTION

10 Most cell surface and soluble molecules that mediate functions such as recognition, adhesion or
binding have evolved from a common evolutionary precursor (i.e., these proteins have structural
homology). A number of molecules outside the immune system that have similar functions are also
derived from this same evolutionary precursor. These molecules are classified as members of the
15 immunoglobulin (Ig) superfamily. The criteria for a protein to be a member of the Ig superfamily is to
have one or more Ig domains, which are regions of 70-110 amino acid residues in length homologous
to either Ig variable-like (V) or Ig constant-like (C) domains. Members of the Ig superfamily include
antibodies (Ab), T cell receptors (TCRs), class I and II major histocompatibility (MHC) proteins, CD2,
CD3, CD4, CD8, poly-Ig receptors, Fc receptors, neural cell-adhesion molecule (NCAM) and platelet-
20 derived growth factor receptor (PDGFR).

Ig domains (V and C) are regions of conserved amino acid residues that give a polypeptide a
globular tertiary structure called an immunoglobulin (or antibody) fold, which consists of two
approximately parallel layers of β -sheets. Conserved cysteine residues form an intrachain disulfide-
bonded loop, 55-75 amino acid residues in length, which connects the two layers of the β -sheets.

25 Each β -sheet has three or four anti-parallel β -strands of 5-10 amino acid residues. Hydrophobic and
hydrophilic interactions of amino acid residues within the β -strands stabilize the Ig fold (hydrophobic
on inward facing amino acid residues and hydrophilic on the amino acid residues in the outward facing
portion of the strands). A V domain consists of a longer polypeptide than a C domain, with an
additional pair of β -strands in the Ig fold.

30 A consistent feature of Ig superfamily genes is that each sequence of an Ig domain is
encoded by a single exon. It is possible that the superfamily evolved from a gene coding for a single
Ig domain involved in mediating cell-cell interactions. New members of the superfamily then arose by
exon and gene duplications. Modern Ig superfamily proteins contain different numbers of V and/or C

WO 02/072794

PCT/US02/09052

domains. Another evolutionary feature of this superfamily is the ability to undergo DNA rearrangements, a unique feature retained by the antigen receptor members of the family.

Many members of the Ig superfamily are integral plasma membrane proteins with extracellular Ig domains. The hydrophobic amino acid residues of their transmembrane domains and their cytoplasmic tails are very diverse, with little or no homology among Ig family members or to known signal-transducing structures. There are exceptions to this general superfamily description. For example, the cytoplasmic tail of PDGFR has tyrosine kinase activity. In addition Thy-1 is a glycoprotein found on thymocytes and T cells. This protein has no cytoplasmic tail, but is instead attached to the plasma membrane by a covalent glycosphosphatidylinositol linkage.

Another common feature of many Ig superfamily proteins is the interactions between Ig domains which are essential for the function of these molecules. Interactions between Ig domains of a multimeric protein can be either homophilic or heterophilic (i.e., between the same or different Ig domains). Antibodies are multimeric proteins which have both homophilic and heterophilic interactions between Ig domains. Pairing of constant regions of heavy chains forms the Fc region of an antibody and pairing of variable regions of light and heavy chains form the antigen binding site of an antibody. Heterophilic interactions also occur between Ig domains of different molecules. These interactions provide adhesion between cells for significant cell-cell interactions in the immune system and in the developing and mature nervous system. (Reviewed in Abbas, A.K. et al. (1991) Cellular and Molecular Immunology, W.B. Saunders Company, Philadelphia, PA, pp.142-145.)

Antibodies

Antibodies are multimeric members of the Ig superfamily which are either expressed on the surface of B-cells or secreted by B-cells into the circulation. Antibodies bind and neutralize foreign antigens in the blood and other extracellular fluids. The prototypical antibody is a tetramer consisting of two identical heavy polypeptide chains (H-chains) and two identical light polypeptide chains (L-chains) interlinked by disulfide bonds. This arrangement confers the characteristic Y-shape to antibody molecules. Antibodies are classified based on their H-chain composition. The five antibody classes, IgA, IgD, IgE, IgG and IgM, are defined by the α , δ , ϵ , γ , and μ H-chain types. There are two types of L-chains, κ and λ , either of which may associate as a pair with any H-chain pair. IgG, the most common class of antibody found in the circulation, is tetrameric, while the other classes of antibodies are generally variants or multimers of this basic structure.

H-chains and L-chains each contain an N-terminal variable region and a C-terminal constant region. The constant region consists of about 110 amino acids in L-chains and about 330 or 440 amino acids in H-chains. The amino acid sequence of the constant region is nearly identical among H- or L-

WO 02/072794

PCT/US02/09052

chains of a particular class. The variable region consists of about 110 amino acids in both H- and L-chains. However, the amino acid sequence of the variable region differs among H- or L-chains of a particular class. Within each H- or L-chain variable region are three hypervariable regions of extensive sequence diversity, each consisting of about 5 to 10 amino acids. In the antibody molecule, the H- and L-chain hypervariable regions come together to form the antigen recognition site. (Reviewed in Alberts, B. et al. (1994) Molecular Biology of the Cell, Garland Publishing, New York, NY, pp. 1206-1213 and 1216-1217.)

Both H-chains and L-chains contain the repeated Ig domains of members of the Ig superfamily. For example, a typical H-chain contains four Ig domains, three of which occur within the constant region and one of which occurs within the variable region and contributes to the formation of the antigen recognition site. Likewise, a typical L-chain contains two Ig domains, one of which occurs within the constant region and one of which occurs within the variable region.

The immune system is capable of recognizing and responding to any foreign molecule that enters the body. Therefore, the immune system must be armed with a full repertoire of antibodies against all potential antigens. Such antibody diversity is generated by somatic rearrangement of gene segments encoding variable and constant regions. These gene segments are joined together by site-specific recombination which occurs between highly conserved DNA sequences that flank each gene segment. Because there are hundreds of different gene segments, millions of unique genes can be generated combinatorially. In addition, imprecise joining of these segments and an unusually high rate of somatic mutation within these segments further contribute to the generation of a diverse antibody population.

Neural Cell Adhesion Proteins

Neural cell adhesion proteins (NCAPs) play roles in the establishment of neural networks during development and regeneration of the nervous system (Uyemura et al. (1996) *Essays Biochem.* 31:37-48; Brummendorf and Rathjen (1996) *Curr. Opin. Neurobiol.* 6:584-593). NCAP participates in neuronal cell migration, cell adhesion, neurite outgrowth, axonal fasciculation, pathfinding, synaptic target-recognition, synaptic formation, myelination and regeneration. NCAPs are expressed on the surfaces of neurons associated with learning and memory. Mutations in genes encoding NCAPs are linked with neurological diseases, including Charcot-Marie-Tooth disease (a hereditary neuropathy), Dejerine-Sottas disease, X-linked hydrocephalus, MASA syndrome (mental retardation, aphasia, shuffling gait and adducted thumbs), and spastic paraplegia type I. In some cases, expression of NCAP is not restricted to the nervous system. L1, for example, is expressed in melanoma cells and hematopoietic tumor cells where it is implicated in cell spreading and migration, and may play a role in

WO 02/072794

PCT/US02/09052

tumor progression (Montgomery et al. (1996) *J. Cell Biol.* 132:475-485).

NCAPs have at least one immunoglobulin constant or variable domain (Uyemura et al., *supra*). They are generally linked to the plasma membrane through a transmembrane domain and/or a glycosyl-phosphatidylinositol (GPI) anchor. The GPI linkage can be cleaved by GPI phospholipase C.

5 Most NCAPs consist of an extracellular region made up of one or more immunoglobulin domains, a membrane spanning domain, and an intracellular region. Many NCAPs contain post-translational modifications including covalently attached oligosaccharide, glucuronic acid, and sulfate. NCAPs fall into three subgroups: simple-type, complex-type, and mixed-type. Simple-type NCAPs contain one or more variable or constant immunoglobulin domains, but lack other types of domains. Members of the

10 simple-type subgroup include Schwann cell myelin protein (SMP), limbic system-associated membrane protein (LAMP) and opiate-binding cell-adhesion molecule (OBCAM). The complex-type NCAPs contain fibronectin type III domains in addition to the immunoglobulin domains. The complex-type subgroup includes neural cell-adhesion molecule (NCAM), axonin-1, F11, Bravo, and L1. Mixed-type NCAPs contain a combination of immunoglobulin domains and other motifs such as tyrosine kinase,

15 epidermal growth factor-like, sema, and PSI (plexins, semaphorins, and integrins) domains. This subgroup includes Trk receptors of nerve growth factors such as nerve growth factor (NGF) and neurotrophin 4 (NT4), Neu differentiation factors such as glial growth factor II (GGFII) and acetylcholine receptor-inducing factor (ARIA), the semaphorin/collapsin family such as semaphorin B and collapsin, and receptors for members of the semaphorin/collapsin family such as plexin (for plexin,

20 see below).

An NCAP subfamily, the NCAP-LON subgroup, includes cell adhesion proteins expressed on distinct subpopulations of brain neurons. Members of the NCAP-LON subgroup possess three immunoglobulin domains and bind to cell membranes through GPI anchors. Kilon (a kindred of NCAP-LON), for example, is expressed in the brain cerebral cortex and hippocampus (Funatsu et al.

25 (1999) *J. Biol. Chem.* 274:8224-8230). Immunostaining localizes Kilon to the dendrites and soma of pyramidal neurons. Kilon has three C2 type immunoglobulin-like domains, six predicted glycosylation sites, and a GPI anchor. Expression of Kilon is developmentally regulated. It is expressed at higher levels in adult brain in comparison to embryonic and early postnatal brains. Confocal microscopy shows the presence of Kilon in dendrites of hypothalamic magnocellular neurons secreting

30 neuropeptides, oxytocin, or arginine vasopressin (Miyata et al. (2000) *J. Comp. Neurol.* 424:74-85). Arginine vasopressin regulates body fluid homeostasis, extracellular osmolarity and intravascular volume. Oxytocin induces contractions of uterine smooth muscle during child birth and of myoepithelial cells in mammary glands during lactation. In magnocellular neurons, Kilon is proposed to

WO 02/072794

PCT/US02/09052

play roles in the reorganization of dendritic connections during neuropeptide secretion.

Sidekick (SDK) is a member of the NCAP family. The extracellular region of SDK contains six immunoglobulin domains and thirteen fibronectin type III domains. SDK is involved in cell-cell interaction during eye development in *Drosophila* (Nguyen, D. N. T. et al. (1997) *Development* 124: 3303).

Synaptic Membrane Glycoproteins

Specialized cell junctions can occur at points of cell-cell contact. Among these cell junctions are communicating junctions which mediate the passage of chemical and electrical signals between cells. In the central nervous system, communicating junctions between neurons are known as synaptic junctions. They are composed of the membranes and cytoskeletons of the pre- and post-synaptic neurons. Some glycoproteins, found in biochemically isolated synaptic subfractions such as the synaptic membrane (SM) and postsynaptic density (PSD) fractions, have been identified and their functions established. An example is the SM glycoprotein, gp50, identified as the $\beta 2$ subunit of the Na⁺/K⁺-ATPase.

Two glycoproteins, gp65 and gp55, are major components of synaptic membranes prepared from rat forebrain. They are members of the Ig superfamily containing three and two Ig domains, respectively. As members of the Ig superfamily, it is proposed that a possible function of these proteins is to mediate adhesive interactions at the synaptic junction. (Langnaese, K. et al. (1997) *J. Biol. Chem.* 272:821-827.)

Lectins

Lectins comprise a ubiquitous family of extracellular glycoproteins which bind cell surface carbohydrates specifically and reversibly, resulting in the agglutination of cells (reviewed in Drickamer, K. and Taylor, M. E. (1993) *Annu. Rev. Cell Biol.* 9:237-264). This function is particularly important for activation of the immune response. Lectins mediate the agglutination and mitogenic stimulation of lymphocytes at sites of inflammation (Lasky, L. A. (1991) *J. Cell. Biochem.* 45:139-146; Palletta, E. et al. (1989) *J. Immunol.* 143:2850-2857).

Sialic acid binding Ig-like lectins (SIGLECs) are members of the Ig superfamily that bind to sialic acids in glycoproteins and glycolipids. SIGLECs include sialoadhesin, CD22, CD33, myelin-associated glycoprotein (MAG), SIGLEC-5, SIGLEC-6, SIGLEC-7, and SIGLEC-8. The extracellular region of SIGLEC has a membrane distal V-set domain followed by varying numbers of C2-set domains. The sialic acid binding domain is mapped to the V-set domain. Except for MAG which is expressed exclusively in the nervous system, most SIGLECs are expressed on distinct subsets of hemopoietic cells. For example, SIGLEC-8 is expressed exclusively in eosinophils, one form of

WO 02/072794

PCT/US02/09052

polymorphonuclear leucocyte (granulocyte) (Floyd, H. et al. (2000) J. Biol. Chem. 275: 861-866).

Leucine-Rich Repeat Proteins

Leucine-rich repeat proteins (LRRs) are involved in protein-protein interactions. LRRs such as mammalian neuronal leucine-rich repeat proteins (NLLR-1 and NLLR-2), *Drosophila* connectin, slit, chaopin, and toll all play roles in neuronal development. The extracellular region of LRRs contains varying numbers of leucine-rich repeats, immunoglobulin-like domains, and fibronectin type III domains (Taguchi, A. et al. (1996) Brain Res. Mol. Brain Res. 35:31-40).

In addition to the V and C2 sets of immunoglobulin-like domains, there is a D set immunoglobulin-like domain, named IPT/TIG (for immunoglobulin-like fold shared by plexins and transcription factors). IPT/TIG containing proteins include plexins, MET/ RON/ SEA (hepatocyte growth factor receptor family), and the transcription factor XCo2, a transcription factor of the Col/Olf-1/EBF family involved in the specification of primary neurons in *Xenopus* (Bork, P. et al. (1999) Trends in Biochem. 24:261-263; Santoro, N. M. et al. (1996) Mol. Cell Biol. 16:7072-7083; Dubois L. et al. (1998) Curr. Biol.8:199-209). Plexins such as plexin A and VESPR have been shown to be neuronal semaphorin receptors that control axon guidance (Winberg M. L. et al. (1998) Cell 95:903-916).

Expression profiling

Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

The discovery of new immunoglobulin superfamily proteins, 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 immune system, neurological, developmental, muscle, and cell proliferative disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of immunoglobulin superfamily proteins.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, immunoglobulin superfamily proteins, referred to collectively as "IGSFP" and individually as "IGSFP-1," "IGSFP-2," "IGSFP-3," "IGSFP-4," "IGSFP-

WO 02/072794

PCT/US02/09052

5," "IGSFP-6," "IGSFP-7," "IGSFP-8," "IGSFP-9," "IGSFP-10," "IGSFP-11," and "IGSFP-12." 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:1-12, 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:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-12.

10 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:1-12, 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:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-12. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:13-24.

15 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:1-12, 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:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. 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.

20 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:1-12, 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:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group

WO 02/072794

PCT/US02/09052

consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

5 Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, 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:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

The invention further provides an isolated polynucleotide selected from the group consisting of
15 a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, 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:13-24, 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
20 comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of
a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least
25 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to
30 said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

WO 02/072794

PCT/US02/09052

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:13-24, 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:13-24, 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:1-12, 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:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional IGSFP, 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:1-12, 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:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. 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 IGSFP, comprising

WO 02/072794

PCT/US02/09052

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:1-12, 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:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. 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 IGSFP, 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:1-12, 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:1-12, c) a biologically active fragment of a polypeptide
20 having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. 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:1-12, 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:1-12, c) a biologically active fragment of a polypeptide
30 having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. 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

WO 02/072794

PCT/US02/09052

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

5 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:13-24, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target
10 polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

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
15 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:13-24, 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:13-24, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of
20 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:13-24, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a
25 polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, 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
30 biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

WO 02/072794

PCT/US02/09052

BRIEF DESCRIPTION OF THE TABLES

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

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are 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

WO 02/072794

PCT/US02/09052

used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

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

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

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

"Altered" nucleic acid sequences encoding IGSFP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as IGSFP or a polypeptide with at least one functional characteristic of IGSFP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding IGSFP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding IGSFP. 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 IGSFP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of IGSFP 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

WO 02/072794

PCT/US02/09052

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

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

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

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

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

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

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

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No.

WO 02/072794

PCT/US02/09052

5.270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH₂), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) *J. Biotechnol.* 74:5-13.)

10 The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) *Proc. Natl Acad. Sci. USA* 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

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

25 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 IGSFP, 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

WO 02/072794

PCT/US02/09052

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 IGSPF or fragments of IGSPF 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
	Ala	Gly, Ser
25	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
30	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
35	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr

WO 02/072794

PCT/US02/09052

Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

5

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.

10 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
 15 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.

20 "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
 25 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 IGSFP or the polynucleotide encoding IGSFP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up
 30 to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For
 35 example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected

WO 02/072794

PCT/US02/09052

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.

5 A fragment of SEQ ID NO:13-24 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:13-24, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:13-24 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:13-24 from related polynucleotide sequences. The precise length of a fragment of SEQ ID
10 NO:13-24 and the region of SEQ ID NO:13-24 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-12 is encoded by a fragment of SEQ ID NO:13-24. A fragment of SEQ ID NO:1-12 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-12. For example, a fragment of SEQ ID NO:1-12 is useful as an
15 immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-12. The precise length of a fragment of SEQ ID NO:1-12 and the region of SEQ ID NO:1-12 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
20 (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

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

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

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

WO 02/072794

PCT/US02/09052

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.

5 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
10 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/b12.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST
15 programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

20 *Penalty for mismatch: -2*

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

25 *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
30 nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

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

WO 02/072794

PCT/US02/09052

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.

5 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.

10 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 15 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 20 example:

```
Matrix: BLOSUM62
Open Gap: 11 and Extension Gap: 1 penalties
Gap x drop-off: 50
Expect: 10
25 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 30 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.

WO 02/072794

PCT/US02/09052

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

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

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

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

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances,

WO 02/072794

PCT/US02/09052

such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

5 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

10 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

15 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 IGSFP 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

20 of IGSFP 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.

25 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 IGSFP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of IGSFP.

30 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

WO 02/072794

PCT/US02/09052

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.

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

10 "Post-translational modification" of an IGSFP 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 IGSFP.

"Probe" refers to nucleic acid sequences encoding IGSFP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are 15 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 20 DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

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

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold 30 Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that

WO 02/072794

PCT/US02/09052

purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 5 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase 10 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 15 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 20 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.

25 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 30 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

WO 02/072794

PCT/US02/09052

vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

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

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

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

The term "sample" is used in its broadest sense. A sample suspected of containing IGSFP, nucleic acids encoding IGSFP, 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,

WO 02/072794

PCT/US02/09052

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

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

A "transgenic organism," as used herein, is any organism, including but not limited to animals
15 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. In one alternative, the nucleic acid can be introduced by infection with a
20 recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). 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
25 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
30 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

WO 02/072794

PCT/US02/09052

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
5 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
10 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
15 the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

20

THE INVENTION

The invention is based on the discovery of new human immunoglobulin superfamily proteins (IGSFP), the polynucleotides encoding IGSFP, and the use of these compositions for the diagnosis, treatment, or prevention of immune system, neurological, developmental, muscle, and cell proliferative
25 disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte
30 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. Column 6 shows the Incyte ID numbers of physical, full length clones corresponding to the polypeptide and

WO 02/072794

PCT/US02/09052

polynucleotide sequences of the invention. The full length clones encode polypeptides which have at least 95% sequence identity to the polypeptide sequences shown in column 3.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database.

5 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 and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches
10 between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

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

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are immunoglobulin superfamily proteins. For example, SEQ ID NO:2 is 50% identical, from residue Q34 to residue P563, to Mus musculus Fca/m
25 receptor (GenBank ID g11071950) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 9.6e-121, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains an immunoglobulin 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.)
30 Data from additional BLAST analyses provide further corroborative evidence that SEQ ID NO:2 is an immunoglobulin. In an alternative example, SEQ ID NO:3 is 40% identical, from residue L30 to residue V176, to surface protein MCA-32 (GenBank ID g1136501) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 6.9e-35, which

WO 02/072794

PCT/US02/09052

indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:3 also contains an immunoglobulin 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 additional BLAST analyses provide further corroborative evidence that SEQ ID NO:3 is a surface protein. In an alternative example, SEQ ID NO:8 is 86% identical, from residue M1 to residue S433, to cell-surface molecule Ly-9 (GenBank ID g10197717) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $7.4e-191$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:8 also contains immunoglobulin domains 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 additional BLAST analysis provide further corroborative evidence that SEQ ID NO:8 is a cell surface molecule which is a member of the immunoglobulin superfamily. In an alternative example, SEQ ID NO:11 is 52% identical, from residue N43 to residue Q604, to human NEPH1 (GenBank ID g14572521) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $5.4e-158$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. As determined by BLAST analysis using the PROTEOME database, SEQ ID NO:11 is localized to the plasma membrane, is homologous to a human protein which contains an immunoglobulin domain and has a region of low similarity to a region of an opioid-binding cell adhesion molecule, which is a glycosylphosphatidylinositol (GPI)-anchored neural cell adhesion molecule (PROTEOME ID 598720|FLJ10845); SEQ ID NO:11 is also homologous to human Nephhrin which is a member of the immunoglobulin superfamily expressed in renal glomeruli which may have a role in the development or function of the kidney filtration barrier. Mutation of the Nephhrin gene causes congenital nephrotic syndrome (PROTEOME ID 340970|NPHS1). SEQ ID NO:11 also contains an immunoglobulin 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 additional BLAST analyses provide further corroborative evidence that SEQ ID NO:11 is a member of the immunoglobulin superfamily. SEQ ID NO:1, SEQ ID NO:4-7, SEQ ID NO:9-10 and SEQ ID NO:12 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-12 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

WO 02/072794

PCT/US02/09052

combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:13-24 or that distinguish between SEQ ID NO:13-24 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 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 polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_XXXXXX_{N₁}_N₂_YYYYY_{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 YYYYY 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 polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FLXXXXXX_gAAAAA_gBBBBB_1_N is 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 02/072794

PCT/US02/09052

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 Table 4 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 IGSFP variants. A preferred IGSFP 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 IGSFP amino acid sequence, and which contains at least one functional or structural characteristic of IGSFP.

The invention also encompasses polynucleotides which encode IGSFP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:13-24, which encodes IGSFP. The polynucleotide sequences of SEQ ID NO:13-24, 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 IGSFP. In

WO 02/072794

PCT/US02/09052

particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding IGSFP. 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:13-
5 24 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:13-24. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of IGSFP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a
10 polynucleotide sequence encoding IGSFP. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding IGSFP, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence
15 identity to the polynucleotide sequence encoding IGSFP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding IGSFP. For example, a polynucleotide comprising a sequence of SEQ ID NO:14 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:24 and a polynucleotide
20 comprising a sequence of SEQ ID NO:16 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:17. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of IGSFP.

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 IGSFP, some bearing minimal
25 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 IGSFP, and all such variations are to be considered as
30 being specifically disclosed.

Although nucleotide sequences which encode IGSFP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring IGSFP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding IGSFP or

WO 02/072794

PCT/US02/09052

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

The invention also encompasses production of DNA sequences which encode IGSFP and IGSFP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the
10 synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding IGSFP 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
15 NO:13-24 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment
20 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),
25 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.
30 (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 IGSFP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences,

WO 02/072794

PCT/US02/09052

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., Trigila, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060.)

Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

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

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire 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.

WO 02/072794

PCT/US02/09052

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode IGSFP may be cloned in recombinant DNA molecules that direct expression of IGSFP, 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
5 equivalent amino acid sequence may be produced and used to express IGSFP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter IGSFP-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
10 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 No.
15 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 IGSFP, 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
20 selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively,
25 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 IGSFP 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
30 Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, IGSFP 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.

WO 02/072794

PCT/US02/09052

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 IGSFP, 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.)

10 In order to express a biologically active IGSFP, the nucleotide sequences encoding IGSFP 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 IGSFP. Such elements may vary in their strength and specificity. Specific initiation signals 15 may also be used to achieve more efficient translation of sequences encoding IGSFP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding IGSFP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may 20 be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.) 25

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding IGSFP 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.) 30

A variety of expression vector/host systems may be utilized to contain and express sequences encoding IGSFP. These include, but are not limited to, microorganisms such as bacteria transformed

WO 02/072794

PCT/US02/09052

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
5 animal cell systems. (See, e.g., Sambrook, *supra*; Ausubel, *supra*; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; *The McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and
10 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)
15 *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding IGSFP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding IGSFP can be achieved using a
20 multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding IGSFP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for
in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of
25 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 IGSFP are needed, e.g. for the production of antibodies, vectors which direct high level expression of IGSFP 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 IGSFP. A number of vectors
30 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*;

WO 02/072794

PCT/US02/09052

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 IGSFP. Transcription of sequences encoding IGSFP 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 IGSFP 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 IGSFP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

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

For long term production of recombinant proteins in mammalian systems, stable expression of IGSFP in cell lines is preferred. For example, sequences encoding IGSFP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

WO 02/072794

PCT/US02/09052

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 *apr* 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
5 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
10 alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β glucuronidase and its substrate β -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system.
15 (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

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

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

Immunological methods for detecting and measuring the expression of IGSFP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques
30 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 IGSFP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See,

WO 02/072794

PCT/US02/09052

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.)

5 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 IGSFP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding IGSFP, or any fragments thereof, may be cloned into a vector
10 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
15 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 IGSFP 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
20 and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode IGSFP may be designed to contain signal sequences which direct secretion of IGSFP 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
25 the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities
30 (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

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

WO 02/072794

PCT/US02/09052

fusion protein in any of the aforementioned host systems. For example, a chimeric IGSFP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of IGSFP 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 IGSFP encoding sequence and the heterologous protein sequence, so that IGSFP 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 IGSFP 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.

IGSFP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to IGSFP. At least one and up to a plurality of test compounds may be screened for specific binding to IGSFP. 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 IGSFP, 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 IGSFP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express IGSFP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing IGSFP or cell membrane fractions which contain IGSFP are then contacted

WO 02/072794

PCT/US02/09052

with a test compound and binding, stimulation, or inhibition of activity of either IGSFP 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 IGSFP, either in solution or affixed to a solid support, and detecting the binding of IGSFP 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.

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

In another embodiment, polynucleotides encoding IGSFP 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 No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) *Science* 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) *Clin. Invest.* 97:1999-2002; Wagner, K.U. et al. (1997) *Nucleic Acids Res.* 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and

WO 02/072794

PCT/US02/09052

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 IGSFP 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 IGSFP 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 IGSFP 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 IGSFP, e.g., by secreting IGSFP 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 IGSFP and immunoglobulin superfamily proteins. In addition, the expression of IGSFP is closely associated with brain, colon, diseased skin, diseased lung, hippocampus, spleen, and diseased vermiform tissues, as well as, CD4⁺ T and peripheral blood cells. Therefore, IGSFP appears to play a role in immune system, neurological, developmental, muscle, and cell proliferative disorders. In the treatment of disorders associated with increased IGSFP expression or activity, it is desirable to decrease the expression or activity of IGSFP. In the treatment of disorders associated with decreased IGSFP expression or activity, it is desirable to increase the expression or activity of IGSFP.

Therefore, in one embodiment, IGSFP 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 IGSFP. Examples of such disorders include, but are not limited to, an immune system disorder such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVID), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies,

WO 02/072794

PCT/US02/09052

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, 5 erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, 10 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 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 15 neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases 20 of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, 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, 25 inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder such as renal tubular acidosis, anemia, 30 Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and

WO 02/072794

PCT/US02/09052

neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline
5 myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, and ethanol myopathy; and 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
10 adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

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

In a further embodiment, a composition comprising a substantially purified IGSFP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of IGSFP including, but not limited to,
20 those provided above.

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

In a further embodiment, an antagonist of IGSFP may be administered to a subject to treat or
25 prevent a disorder associated with increased expression or activity of IGSFP. Examples of such disorders include, but are not limited to, those immune system, neurological, developmental, muscle, and cell proliferative disorders described above. In one aspect, an antibody which specifically binds IGSFP 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 IGSFP.

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

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

WO 02/072794

PCT/US02/09052

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

An antagonist of IGSFP may be produced using methods which are generally known in the art. In particular, purified IGSFP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind IGSFP. Antibodies to IGSFP 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. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

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

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to IGSFP 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 IGSFP 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 IGSFP 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

WO 02/072794

PCT/US02/09052

technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 6:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the
5 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
10 chain antibodies may be adapted, using methods known in the art, to produce IGSFP-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
15 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 IGSFP may also be generated. For example, such fragments include, but are not limited to, F(ab'), fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of
20 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
25 polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between IGSFP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering IGSFP epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

30 Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for IGSFP. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of IGSFP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a

WO 02/072794

PCT/US02/09052

determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple IGSFP epitopes, represents the average affinity, or avidity, of the antibodies for IGSFP. The K_d determined for a preparation of monoclonal antibodies, which are monospecific for a particular IGSFP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_d ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the IGSFP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_d ranging from about 10^6 to 10^9 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of IGSFP, 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 IGSFP-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 IGSFP, 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 IGSFP. 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 IGSFP. (See, e.g., Agrawal, S., ed. (1996) *Antisense Therapeutics*, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, *supra*; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other

WO 02/072794

PCT/US02/09052

gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

5 In another embodiment of the invention, polynucleotides encoding IGSFP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), 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) 10 express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides 20 brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in IGSFP expression or regulation causes disease, the expression of IGSFP 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 25 IGSFP are treated by constructing mammalian expression vectors encoding IGSFP and introducing these vectors by mechanical means into IGSFP-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J.-L. and H. Récapon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of IGSFP include, but are not limited to, the pCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors

WO 02/072794

PCT/US02/09052

(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). IGSFP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter 5 (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous 10 gene encoding IGSFP 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 15 polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with 20 respect to IGSFP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding IGSFP 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 required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are 25 commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and 30 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 No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference.

WO 02/072794

PCT/US02/09052

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).

5 In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding IGSFP to cells which have one or more genetic abnormalities with respect to the expression of IGSFP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 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.

10 In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding IGSFP to target cells which have one or more genetic abnormalities with respect to the expression of IGSFP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing IGSFP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 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 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

WO 02/072794

PCT/US02/09052

ordinary skill in the art.

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

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

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

WO 02/072794

PCT/US02/09052

endonucleolytic cleavage of sequences encoding IGSFP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

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

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

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

WO 02/072794

PCT/US02/09052

decreased IGSFP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding IGSFP 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 IGSFP 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 IGSFP 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 IGSFP. 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 02/072794

PCT/US02/09052

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.

5 Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of IGSFP, antibodies to IGSFP, and mimetics, agonists, antagonists, or inhibitors of IGSFP.

10 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
20 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
25 macromolecules comprising IGSFP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, IGSFP 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).
30

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

WO 02/072794

PCT/US02/09052

route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example IGSFP or fragments thereof, antibodies of IGSFP, and agonists, antagonists or inhibitors of IGSFP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED₅₀ (the dose therapeutically effective in 50% of the population) or LD₅₀ (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD₅₀/ED₅₀ ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED₅₀ with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

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

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

DIAGNOSTICS

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

WO 02/072794

PCT/US02/09052

body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

5 A variety of protocols for measuring IGSFP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of IGSFP expression. Normal or standard values for IGSFP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to IGSFP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of IGSFP expressed in
10 subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding IGSFP 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
15 and quantify gene expression in biopsied tissues in which expression of IGSFP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of IGSFP, and to monitor regulation of IGSFP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding IGSFP or closely related molecules may be used to
20 identify nucleic acid sequences which encode IGSFP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding IGSFP, allelic variants, or related sequences.

25 Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the IGSFP 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:13-24 or from genomic sequences including promoters, enhancers, and introns of the IGSFP gene.

Means for producing specific hybridization probes for DNAs encoding IGSFP include the
30 cloning of polynucleotide sequences encoding IGSFP or IGSFP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a

WO 02/072794

PCT/US02/09052

variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding IGSFP may be used for the diagnosis of disorders associated with expression of IGSFP. Examples of such disorders include, but are not limited to, an

5 immune system disorder such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobulinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema,

10 immunodeficiency associated with Cushing's disease, 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,

15 erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis,

20 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 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

25 neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases

30 of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other

WO 02/072794

PCT/US02/09052

neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, 5 postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, 10 hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline 15 myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, and ethanol myopathy; and 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 20 adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding IGSFP may be used in Southern or northern analysis, dot blot, or other 25 membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered IGSFP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding IGSFP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide 30 sequences encoding IGSFP 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

WO 02/072794

PCT/US02/09052

sample then the presence of altered levels of nucleotide sequences encoding IGSFP 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.

5 In order to provide a basis for the diagnosis of a disorder associated with expression of IGSFP, 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 IGSFP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal
10 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,
15 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
20 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.

25 Additional diagnostic uses for oligonucleotides designed from the sequences encoding IGSFP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding IGSFP, or a fragment of a polynucleotide complementary to the polynucleotide encoding IGSFP, and will be employed under optimized conditions for identification of a specific gene or
30 condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding IGSFP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are

WO 02/072794

PCT/US02/09052

substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding IGSFP are used to amplify DNA using the
5 polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as
10 DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the
15 alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis,
20 sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the
25 anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations. (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu
30 (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641.)

Methods which may also be used to quantify the expression of IGSFP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C.

WO 02/072794

PCT/US02/09052

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.

5 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
10 function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and
15 display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

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

20 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 No.
25 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
30 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.

WO 02/072794

PCT/US02/09052

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

5 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

10 from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a

15 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

20 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

25 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

30 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

WO 02/072794

PCT/US02/09052

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
5 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
10 spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein
15 identification.

A proteomic profile may also be generated using antibodies specific for IGSFP to quantify the levels of IGSFP 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-
20 111; Mendozze, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and
25 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
30 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

WO 02/072794

PCT/US02/09052

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.

5 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
10 with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA
15 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

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

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic

WO 02/072794

PCT/US02/09052

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 IGSFP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the
5 region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is
10 valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant
15 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, IGSFP, 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
20 solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between IGSFP 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
25 synthesized on a solid substrate. The test compounds are reacted with IGSFP, or fragments thereof, and washed. Bound IGSFP is then detected by methods well known in the art. Purified IGSFP 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.

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

WO 02/072794

PCT/US02/09052

In additional embodiments, the nucleotide sequences which encode IGSFP 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.

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

The disclosures of all patents, applications and publications, mentioned above and below, 10 including U.S. Ser. No. 60/275,249, U.S. Ser. No. 60/316,810, U.S. Ser. No. 60/323,977, U.S. Ser. No. 60/348,447, and U.S. Ser. No. 60/343,880, are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

15 Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). 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 20 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 25 (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 30 UNIZAP vector system (Stratagene) or SUPERSKRIP 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

WO 02/072794

PCT/US02/09052

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.,

- 5 PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from
10 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
15 AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

- Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a
20 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).

25 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 cyclor or the PTC-200 thermal cyclor (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the
30 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

WO 02/072794

PCT/US02/09052

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, 5 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 10 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; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); hidden Markov 15 model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) *Nucleic Acids Res.* 29:41-43); and HMM-based protein domain databases such as SMART (Schultz et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:5857-5864; Letunic, I. et al. (2002) *Nucleic Acids Res.* 30:242-244). (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.) 20 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 25 Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein 30 databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software

WO 02/072794

PCT/US02/09052

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

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

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:13-24. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative immunoglobulin superfamily proteins were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg).

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

WO 02/072794

PCT/US02/09052

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

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

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

"Stretched" Sequences

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

WO 02/072794

PCT/US02/09052

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 IGSFP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:13-24 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:13-24 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.)

Analogous computer techniques applying BLAST were used to search for identical or related

WO 02/072794

PCT/US02/09052

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:

5

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

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 IGSFP 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

WO 02/072794

PCT/US02/09052

by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding IGSFP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of IGSFP Encoding Polynucleotides

5 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
10 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.

15 High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+} , $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters
20 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.

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

The extended nucleotides were desalted and concentrated, transferred to 384-well plates,

WO 02/072794

PCT/US02/09052

digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (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. Identification of Single Nucleotide Polymorphisms in IGSFP Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:13-24 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase,

WO 02/072794

PCT/US02/09052

polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

5 Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezuelan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The
10 Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

15 X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:13-24 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
20 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ - 32 P] 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 10^7 counts per minute of the labeled probe is used in a typical membrane-based
25 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
30 under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

WO 02/072794

PCT/US02/09052

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).

5 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*
10 270:467-470; 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
15 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
20 complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and
25 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
30 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 02/072794

PCT/US02/09052

using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended
5 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 washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR
15 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 U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average
20 concentration of 100 ng/ μ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

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

Hybridization

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

WO 02/072794

PCT/US02/09052

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

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an
5 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
10 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
15 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
20 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
25 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
30 signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and
measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission
spectra) between the fluorophores using each fluorophore's emission spectrum.

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

WO 02/072794

PCT/US02/09052

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).

For example, SEQ ID NO:19 showed differential expression in toxicology studies as determined by microarray analysis. The expression of SEQ ID NO:19 was decreased by at least two fold in a human C3A liver cell line treated with various drugs (e.g., steroids, steroid hormones) relative to untreated C3A cells. The human C3A cell line is a clonal derivative of HepG2/C3 (hepatoma cell line, isolated from a 15-year-old male with liver tumor), which was selected for strong contact inhibition of growth. The C3A cell line is well established as an *in vitro* model of the mature human liver (Mickelson et al. (1995) *Hepatology* 22:866-875; Nagendra et al. (1997) *Am J Physiol* 272:G408-G416). Effects upon liver metabolism are important to understanding the pharmacodynamics of a drug. Therefore, SEQ ID NO:19 is useful for understanding the pharmacodynamics of a drug.

XII. Complementary Polynucleotides

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

XIII. Expression of IGSFP

Expression and purification of IGSFP is achieved using bacterial or virus-based expression systems. For expression of IGSFP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express IGSFP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of IGSFP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding IGSFP by either homologous recombination or bacterial-mediated

WO 02/072794

PCT/US02/09052

transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, IGSFP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from IGSFP 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 IGSFP obtained by these methods can be used directly in the assays shown in Examples XVII and XVIII where applicable.

XIV. Functional Assays

IGSFP function is assessed by expressing the sequences encoding IGSFP 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 μg 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 μg 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

WO 02/072794

PCT/US02/09052

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

XV. Production of IGSFP Specific Antibodies

IGSFP 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 animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the IGSFP amino acid sequence is analyzed using LASERGENE software
(DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is
synthesized and used to raise antibodies by means known to those of skill in the art. Methods for
selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well
described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A
peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-
Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to
increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the
oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for
30 antipeptide and anti-IGSFP activity by, for example, binding the peptide or IGSFP to a substrate,
blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat
anti-rabbit IgG.

WO 02/072794

PCT/US02/09052

XVI. Purification of Naturally Occurring IGSFP Using Specific Antibodies

Naturally occurring or recombinant IGSFP is substantially purified by immunoaffinity chromatography using antibodies specific for IGSFP. An immunoaffinity column is constructed by covalently coupling anti-IGSFP antibody to an activated chromatographic resin, such as

5 CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing IGSFP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of IGSFP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt

10 antibody/IGSFP 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 IGSFP is collected.

XVII. Identification of Molecules Which Interact with IGSFP

IGSFP, 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
15 previously arrayed in the wells of a multi-well plate are incubated with the labeled IGSFP, washed, and any wells with labeled IGSFP complex are assayed. Data obtained using different concentrations of IGSFP are used to calculate values for the number, affinity, and association of IGSFP with the candidate molecules.

Alternatively, molecules interacting with IGSFP are analyzed using the yeast two-hybrid
20 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).

IGSFP 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.

25 Patent No. 6,057,101).

XVIII. Demonstration of IGSFP Activity

An assay for IGSFP activity measures the ability of IGSFP to recognize and precipitate antigens from serum. This activity can be measured by the quantitative precipitin reaction. (Golub, E.
S. et al. (1987) *Immunology: A Synthesis*, Sinauer Associates, Sunderland, MA, pages 113-115.)

30 IGSFP is isotopically labeled using methods known in the art. Various serum concentrations are added to constant amounts of labeled IGSFP. IGSFP-antigen complexes precipitate out of solution and are collected by centrifugation. The amount of precipitable IGSFP-antigen complex is proportional to the amount of radioisotope detected in the precipitate. The amount of precipitable

WO 02/072794

PCT/US02/09052

IGSFp-antigen complex is plotted against the serum concentration. For various serum concentrations, a characteristic precipitin curve is obtained, in which the amount of precipitable IGSFP-antigen complex initially increases proportionately with increasing serum concentration, peaks at the equivalence point, and then decreases proportionately with further increases in serum concentration.

5 Thus, the amount of precipitable IGSFP-antigen complex is a measure of IGSFP activity which is characterized by sensitivity to both limiting and excess quantities of antigen.

Alternatively, an assay for IGSFP activity measures the expression of IGSFP on the cell surface. cDNA encoding IGSFP is transfected into a non-leukocytic cell line. Cell surface proteins are labeled with biotin (de la Fuente, M.A. et.al. (1997) *Blood* 90:2398-2405). Immunoprecipitations
10 are performed using IGSFP-specific antibodies, and immunoprecipitated samples are analyzed using SDS-PAGE and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled immunoprecipitant is proportional to the amount of IGSFP expressed on the cell surface.

Alternatively, an assay for IGSFP activity measures the amount of cell aggregation induced by overexpression of IGSFP. In this assay, cultured cells such as NIH3T3 are transfected with cDNA
15 encoding IGSFP contained within a suitable mammalian expression vector under control of a strong promoter. Cotransfection with cDNA encoding a fluorescent marker protein, such as Green Fluorescent Protein (CLONTECH), is useful for identifying stable transfectants. The amount of cell agglutination, or clumping, associated with transfected cells is compared with that associated with untransfected cells. The amount of cell agglutination is a direct measure of IGSFP activity.

20 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.

25 Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	CA2 Reagents
385123	1	385123CD1	13	385123CB1	
4547188	2	4547188CD1	14	4547188CB1	9006591CA2
3939883	3	3939883CD1	15	3939883CB1	
3163819	4	3163819CD1	16	3163819CB1	3163819CA2
8518269	5	8518269CD1	17	8518269CB1	90110559CA2
1592646	6	1592646CD1	18	1592646CB1	
7500191	7	7500191CD1	19	7500191CB1	
7500089	8	7500089CD1	20	7500089CB1	2836421CA2
7682434	9	7682434CD1	21	7682434CB1	
2202389	10	2202389CD1	22	2202389CB1	
7503597	11	7503597CD1	23	7503597CB1	
7503603	12	7503603CD1	24	7503603CB1	

WO 02/072794

PCT/US02/09052

Table 2

Polypeptide SEQ ID NO:	Incycle Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	3855123CD1	g14572521	2.00E-70	[Homo sapiens] NEPH1 (D0404834) FcγRIIb receptor (Shibuya, A. et al. (2001) Mol. Cell Biol. 21 (14), 4829-4836)
2	4547188CD1	g11071950	9.60E-121	[Mus musculus] (A3048834) FcγRIIb receptor (Shibuya, A. et al. (2001) Nat. Immunol. 1 (5), 441-446)
3	3929883CD1	g1136501	6.90E-35	[Rattus norvegicus] surface protein MCA-32 (Pozzati, G. et al. (1995) J. Immunol. 155 (12), 5811-5818)
4	3163819CD1	g9887089	6.50E-32	[Mus musculus] lymphocyte antigen 108 isoform 1 (Peck, S. R. et al. (2000) Immunogenetics 52 (1-2), 63-72)
4	3163819CD1	g15384841	1.00E-112	[Homo sapiens] activating NK receptor (Bottino, C. et al. (2001) The Journal of experimental medicine. 194 (3), 235-246)
5	8518269CD1	g9887089	5.20E-62	[Mus musculus] lymphocyte antigen 108 isoform 1 (Peck, S. R. et al. (2000) Immunogenetics 52 (1-2), 63-72)
5	8518269CD1	g15384841	0.00E+00	[Homo sapiens] activating NK receptor (Bottino, C. et al. (2001) The Journal of experimental medicine. 194 (3), 235-246)
6	1592646CD1	g18376829	1.00E-154	[Homo sapiens] (AF39163) osteoclast-associated receptor OSCAR-M2 (Kim, N. et al. (2002) J. Exp. Med. 195 (2), 201-209)
6	1592646CD1	g2645890	2.00E-32	[Homo sapiens] GCSF1 (Mazzarella, R. et al. (1998) Genomics 48 (2), 157-162)
7	7500191CD1	g2078518	0	[Homo sapiens] hecogenin (Vielmeier, J. et al. (1997) Genomics 41 (3), 414-421)
8	7500099CD1	g10197717	7.40E-191	[Homo sapiens] cell-surface molecule Ly-9 (Tovar, V. et al. (2000) Immunogenetics 51 (10), 788-793)
9	7682434CD1	g586	1.20E-80	[Bos taurus] put. pre-OPCAM (AA 1 - 345) (Schofield, P. R. et al. (1989) EMBO J. 8 (2), 489-495)

WO 02/072794

PCT/US02/09052

Table 2

Polypeptide SEQ ID NO:	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	3346698[OPCWIL]	2.2E-81	[Homo sapiens][Receptor (signaling)][Plasma membrane] Opioid-binding cell adhesion molecule, has strong similarity to ratRn.11366, which is a glycosylphosphatidylinositol (GPI)-anchored neutral cell adhesion molecule and a member of the immunoglobulin superfamily (Struyk, A. F., et al. (1995) Cloning of neurotrophin defines a new subfamily of differentially expressed neutral cell adhesion molecules. J. Neurosci. 15:2141-2156; Lane, C. M. et al. (1992) Regulation of an opioid-binding protein in NG108-15 cells parallels regulation of delta-opioid receptors. Proc. Natl. Acad. Sci. U.S.A. 89:11234-11238.)
	332056[Rn.11366]	5.3E-80	[Rattus norvegicus][Receptor (signaling)][Plasma membrane] Opioid-binding cell adhesion molecule, member of the immunoglobulin superfamily and a glycosylphosphatidylinositol (GPI)-anchored neutral cell adhesion molecule (Hachisaka, A., et al. (2000) Developmental expression of opioid-binding cell adhesion molecule (OBCAM) in rat brain. Brain Res. Dev. Brain Res. 122:183-191.)
	330088[Lsmp]	3.9E-77	[Rattus norvegicus][Plasma membrane] Limbic system-associated membrane protein, a member of the Ig family of proteins that plays a role in the selective growth of neurons and the targeting of axons (Pimenta, A. F., et al. (1996) cDNA cloning and structural analysis of the human limbic system-associated membrane protein (LAMP). Gene. 170:189-195.)
11	750559[CD1g14572521]	5.40E-158	[Homo sapiens] NEPH1 (Donoviel, D.B. et al. (2001) Mol. Cell. Biol. 21 (14), 4829-4836)

WO 02/072794

PCT/US02/09052

Table 2

Polypeptide SEQ ID NO:	In vivo Polypeptide ID	GenBank ID NO. or PROTEOME ID NO.	Probability Score	Annotation
		598720 FJ10845	6.6E-66	[Homo sapiens] Protein containing an immunoglobulin (Ig) domain, has a region of low similarity to a region of rat Rn.1.366, opiod-binding cell adhesion molecule, which is a glycosylphosphatidylinositol (GPI)-anchored neural cell adhesion molecule
		340970 NPHS1	5.6E-21	[Homo sapiens] Plasma membrane Cell junction Nephrin, a member of the immunoglobulin family expressed in renal glomeruli, may have a role in the development or function of the kidney filtration barrier; mutation of corresponding gene causes congenital nephrotic syndrome (Ruotsalainen, V. et al. (2000) Role of nephrin in cell junction formation in human nephrogenesis. Am. J. Pathol. 157:1905-1916.)

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	3855123CD1	442	S37 S51 S118 S129 S138 S171 S227 S236 S252 S366 S379 S385 S398 T48 T261 T306 T389 Y60	N162	Signal Peptide: M198-C223 Immunoglobulin domain: G13-A64, G97-A165 Transmembrane domain: A193-A221 N-terminus is non-cytosolic IMMUNOGLOBULIN_DM00001 Q08180 426-518; V80-D172	HMMER
2	4547188CD1	577	S39 S108 S189 S206 S301 S405 S482 S493 S525 T6 T38 T88 T234 T360 T271 T335 T349 T350 T437 T486 T524 T569 Y24	N212 N321	Signal Peptide: M46-P63, M46-Q64, P33-P63 Immunoglobulin domain: G120-L200 Transmembrane domain: S39-P67 R495-R517 N-terminus is cytosolic IMMUNOGLOBULIN_DM00001 P01833 41-120; H128-G201 IMMUNOGLOBULIN_DM00001 P15083 41-120; H128-F208 IMMUNOGLOBULIN_DM00001 P01832 28-125; G120-G201	HMMER TMAP BLAST_DOMO BLAST_DOMO BLAST_DOMO

WO 02/072794

PCT/US02/09052

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	3939883CD1	385	S4 S21 S99 S133 S214 S330 S373 T40 T60 T116 T162 T179 T181 T201 T226 T259 T296 T311 T342 Y236 Y335	N93 N102 N131 N193 N199 N224	IMMUNOGLOBULIN DM00001[54864]41-120: HL28-G201 signal_Cleavage: M1-T38	BLAST_DOMO SPSCAN
					Signal Peptide: M1-G41 Intracellular domains: M1-K19, K293-F385 Transmembrane domains: F20-S39, L270-F292 Extracellular domain: T40-K269 Immunoglobulin domain: G91-A147, D182-A240 Receptor Fe immunoglobulin PD01270: T135-V171, R183-P211 P value < 1.3e-3 SURFACE PROTEIN MCA32 PD095298: L30-V164	HMMER TMHMMER TMHMMER HMMER PFAM BLIMPS_PRODUM BLAST_PRODUM
					PLATELET ENDOTHELIAL CELL ADHESION PRECURSOR SIGNAL MOLECULE PECAM1 CD31 ANTIGEN PDI50932: C68-P305 Leucine zipper pattern: L270-L291 Cell attachment sequences: R308-D310 Signal Peptides: M1-G15, M1-L19, M1-N21	BLAST_PRODUM MOTIFS MOTIFS HMMER
4	3163819CD1	221	S43 S52 S76 S143 S157 S180 T148 T188 T215 Y82	N26 N33 N50 N67 N92 N170 N192 N202	Extracellular domain: M1-K114 Transmembrane domain: M1J5-L137 Intracellular domain: R138-V221	TMHMMER TMHMMER TMHMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	7500019ICD1	1450	S46 S64 S81 S156 S294 S451 S606 S620 S677 S731 S834 S939 S1087 S1137 S1203 S1281 S1283 S1291 S1327 S1328 S1385 S1407 S1423 T143 T212 T279 T311 T365 T371 T438 T532 T581 T603 T628 T759 T784 T808 T869 T873 T892 T924 T948 T1051	N73 N210 N326 N470 N489 N639 N715 N909 N1135 N1287	signal_cleavage: M1-A33	SPSCAN
		T1117 T1121 T1187 T1414 Y127 Y408 Y890			Signal Peptides: M1-G30, M1-A33	HMMER
					Fibronectin type III domain: P539-T621, P633-T721, P954-S1044, P439-S525, P739-L821, P853-S942	HMMER_PFAM
					Immunoglobulin domain: G263-A322, G166-V223, G67-A131, S355-A412	HMMER_PFAM
					Cytosolic domain: T1117-A1450	TMEMMER
					Transmembrane domain: L1094-C1116	
					Non-cytosolic domain: M1-M1093	
					Receptor tyrosine kinase class Y proteins BL00790: V450-P476, Y477-G520, S534-K579	BLIMPS_BLOCKS

WO 02/072794

PCT/US02/09052

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	7500099CD1	S51	S6 S17 S46 S128 S163 S179 S229 S316 S321 S400 S431 S453 S512 S524 T73 T122 T141 T142 T160 T192 T212 T252 T277 T438 T439 T487 Y335	N68 N95 N120 N169 N173 N285 N436	Fibronectin type III repeat signature PR00014: T752-P761, A908-Y926, Y1028-P1042 TUMOR SUPPRESSOR NEEGENIN PROTEIN DCC PRECURSOR COLORECTAL GLYCOPROTEIN IMMUNOGLOBULIN FOLD PD041287: D1169-T1448 PD069999: C1116-P1172 NEEGENIN PROTEIN PD20198: M1-R66 TUMOR SUPPRESSOR PD17136: E58-V133 IMMUNOGLOBULIN DM00001P43146R28-410: P341-Q420 DM00001P43146R42-127: F55-I140 FIBRONECTIN TYPE III REPEAT DM00007P43146R935-1014: A945-D1025 DM00007P43146R834-912: T846-N923 Tooth-dependent receptor proteins signature 1: M1-R5 signal_cleavage: M1-G47	BLIMPS_PRINTS BLAST_PRODOM BLAST_DOMO BLAST_DOMO MOTIFS SPSCAN
					Immunoglobulin domain: S171-A224, G60-I133	HMMER_PPAM

WO 02/072794

PCT/US02/09052

Table 3

SEQ ID NO:	Incyte Peptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	7682434CD1	336	S37 S175 S203 S207 S225 S282 S303 T43 T91 T143 T165 T219 T269 T290	N41 N49 N67 N137 N280 N288	Cytosolic domain: K387-T551 Transmembrane domain: L365-W386 Non-cytosolic domain: M1-K364 ANTIGEN LY9 PRECURSOR SIGNAL TRANSMEMBRANE GLYCOPROTEIN IMMUNOGLOBULIN FOLD PD126134; P359-P545 ANTIGEN PRECURSOR SIGNAL IMMUNOGLOBULIN FOLD GLYCOPROTEIN TCCELL SURFACE CD2 TRANSMEMBRANE PD010953; V55-T243 IMMUNOGLOBULIN DM00001[Q01965]I39-210; M161-S232 B-CELL SURFACE GLYCOPROTEIN/BLAST-1 DM03635IP10232I1-239; V31-S232 DM03635IP18181I1-239; L32-S232 signal_cleavage: M1-S30	TMHMMER BLAST_PRODOM BLAST_PRODOM BLAST_DOMO BLAST_DOMO
					Signal Peptides: M1-R26, M1-S30 Immunoglobulin domain: G231-A293, G47-F114, G147-T107 PRECURSOR SIGNAL GLYCOPROTEIN IMMUNOGLOBULIN FOLD CELL ADHESION GPI-ANCHOR PROTEIN ALTERNATIVE PD005605: F35-Q124	HMMER HMMER_FFAM BLAST_PRODOM

Table 3

SEQ ID NO:	Protein ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10	2202389CD1	241	S44 S88 S112 S163 T10 T134 Y39	N75 N94 N110 N213	IMMUNOGLOBULIN DM00000 PF32736139-125: D40-T123 DM00000 PF32736139-212: V136-D206 DM00000 PF32736226-306: I220-A302 signal_cleavage: M1-I25	BLAST_DOMO SPSCAN
11	75033897CD1	766	S159 S207 S215 S272 S373 S387 S454 S465 S474 S507 S551 S560 S576 S690 S703 S709 S722 T230 T301 T384 T585 T630 T713 Y48 Y307 Y396	N167 N253 N324 N498	Immunoglobulin domain: G51-A117 Signal Peptides: M1-E19, M1-Q21, M1-Q23, M1-L22 Immunoglobulin domain: G62-A129, G163-A229, G433-A501, D264-V316, G349-A400 Cytosolic domain: C547-Y766 Transmembrane domain: Y524-E546 Non-cytosolic domain: M1-A523 GLYCOPROTEIN ANTIGEN PRECURSOR PD02327: L141-I152, T169-I190 IRREGULAR CHIASM CROUCHEST PROTEIN PRECURSOR IRREC TRANSMEMBRANE IMMUNOGLOBULIN FOLD GLYCOPROTEIN SIGNAL_CELL ADHESION PD124347: F50-V256, V261-E315	HMMER_Pfam HMMER HMMER_Pfam TMHMMER BLIMPS_PRODUM BLAST_PRODUM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12	7503603CD1	T6 Y24			IMMUNOGLOBULIN DV00001 Q0818031 26: S51-T142 Leucine_Zipper: L8-L29	BLAST_DOMO MOTIFS

Table 4

Polynucleotide SEQ ID NO./ In-cyte ID/ Sequence Length	Sequence Fragments
13/3855123CBI/ 2691	1-751, 214-969, 282-871, 354-864, 370-934, 388-852, 390-848, 417-958, 459-874, 496-1309, 560-1313, 575-1092, 609-1298, 662-1207, 662-1300, 671-1317, 729-1212, 731-1313, 805-1157, 884-1300, 915-1566, 916-1566, 1010-1503, 1018-1306, 1038-1273, 1038-1641, 1070-1566, 1143-1524, 1231-1791, 1280-1566, 1499-2164, 1559-2109, 1559-2220, 1744-1982, 1752-2366, 1770-2621, 1785-2365, 1785-2576, 1815-2098, 1818-2268, 1826-2440, 1846-2604, 1892-2548, 1898-2664, 1938-2234, 1958-2453, 1969-2691, 1976-2621, 1977-2691, 1985-2631, 1992-2564, 1993-2691, 2014-2625, 2017-2625, 2039-2625, 2081-2689, 2104-2665, 2109-2444, 2140-2691, 2175-2561, 2200-2691, 2217-2691, 2233-2677, 2309-2691, 2338-2691, 2360-2675, 2444-2691, 2505-2691
14/4547188CBI/ 2518	1-148, 1-606, 1-762, 40-275, 147-553, 533-897, 736-1036, 736-1069, 861-1483, 870-1149, 879-1056, 900-1427, 920-1229, 923-1356, 1020-1643, 1049-1312, 1049-1654, 1110-1537, 1114-1678, 1145-1500, 1152-1687, 1156-1687, 1190-1458, 1190-1836, 1199-1828, 1200-1793, 1222-1768, 1251-1817, 1300-1453, 1321-1944, 1321-1966, 1344-1935, 1360-2011, 1364-2011, 1397-1864, 1420-1930, 1422-2081, 1429-2001, 1438-1864, 1470-1522, 1487-1522, 1496-2172, 1500-1552, 1516-1697, 1544-2056, 1596-2090, 1597-2204, 1599-1803, 1602-2252, 1603-2197, 1642-2219, 1669-2081, 1686-2215, 1722-2336, 1767-2432, 1786-2448, 1813-2313, 1827-2479, 1833-2266, 1850-2021, 1863-2465, 2015-2210, 2015-2497, 2015-2499, 2034-2315, 2034-2507, 2034-2518, 2045-2430, 2162-2378
15/3939883CBI/ 1522	1-274, 1-467, 71-510, 124-727, 124-753, 124-794, 124-827, 124-892, 140-599, 180-982, 264-821, 274-529, 496-799, 496-954, 517-1052, 602-1075, 613-727, 717-1177, 796-887, 801-1307, 801-1485, 975-1231, 992-1522, 995-1480, 1039-1498, 1073-1522
16/3163819CBI/ 1463	1-287, 1-996, 1-502, 1-646, 1-650, 1-660, 1-970, 67-983, 97-694, 187-1084, 470-886, 1-511, 1-804, 17-844, 27-305, 27-511, 38-511, 43-511, 63-511, 63-759, 119-511, 146-511, 147-511, 446-1361, 476-1073, 566-1463
18/1592646CBI/ 1557	1-758, 40-1554, 182-563, 183-738, 277-841, 360-876, 360-877, 494-694, 714-980, 853-1459, 863-1394, 866-1281, 900-1178, 909-1454, 937-1489, 949-1225, 960-1203

Table 4

Polynucleotide SEQ ID NO/ nucleotide Sequence Length	Sequence Fragments
19/7500191CBI/5553	960-1456, 964-1245, 964-1250, 990-1552, 992-1221, 992-1531, 995-1253, 996-1535, 1010-1522, 1025-1527, 1089-1381, 1091-1310, 1119-1345, 1119-1528, 1119-1546, 1138-1355, 1140-1410, 1140-1465, 1140-1553, 1150-1555, 1158-1391, 1269-1557
	1-500, 30-501, 214-734, 216-618, 216-733, 216-768, 216-815, 216-877, 216-901, 219-765, 226-593, 226-617, 226-698, 228-759, 278-587, 278-697, 278-740, 282-728, 282-729, 308-857, 316-876, 317-434, 455-988, 611-1167, 611-1202, 656-1263, 671-992, 683-1210, 738-1018, 1068-1319, 1127-5470, 1171-1303, 1193-1700, 1214-830, 1223-1830, 1255-1915, 1333-1818, 1343-1625, 1438-2117, 1439-1910, 1450-1914, 1465-1910, 1493-1630, 1509-1910, 1515-2242, 1552-2101, 1606-2156, 1670-1782, 1738-2363, 1780-2313, 1814-2267, 1839-1978, 1866-2453, 1895-2489, 1910-2565, 1942-2600, 2049-2596, 2114-2693, 2243-2484, 2421-2762, 2453-2637, 2665-3303, 2722-3012, 2731-3272, 2735-2970, 2778-3352, 2798-3240, 2819-3125, 2910-3495, 2971-3570, 3001-3281, 3050-3629, 3112-3767, 3147-3428, 3147-3715, 3201-3446, 3201-3568, 3201-3603, 3201-3698, 3201-3763, 3205-3659, 3239-3473, 3280-3894, 3289-3895, 3289-3949, 3419-4085, 3464-3693, 3476-4043, 3491-3784, 3492-4014, 3506-3987, 3546-4147, 3611-4177, 3620-4185, 3628-3868, 3656-4218, 3679-3889, 3679-4107, 3681-4028, 3690-4093, 3714-4354, 3719-4240, 3726-4245, 3773-4022, 3784-4262, 3797-4056, 3798-4065, 3855-3979, 3872-4089, 3889-4393, 3930-4543, 3947-4186, 3998-4654, 4055-4407, 4057-4201, 4017-4613, 4033-4246, 4033-4563, 4053-4417, 4053-4441, 4059-4716, 4065-4531, 4066-4572, 4073-4530, 4081-4419, 4088-4496, 4089-4678, 4148-4688, 4150-4681, 4195-4420, 4195-4431, 4195-4478, 4195-4708, 4195-4738, 4195-4828, 4204-4515, 4219-4809, 4229-4468, 4245-4745, 4245-4888, 4252-4878, 4255-4507, 4256-4500, 4274-4491, 4280-5057, 4281-4833, 4306-4866, 4314-4846, 4330-4828, 4330-4899, 4336-4805, 4358-4635, 4425-4880, 4426-4709, 4429-4770, 4429-4774, 4437-4952, 4444-4665, 4445-5152, 4506-5025, 4512-5096, 4590-5188, 4599-5150, 4602-5114, 4618-5153, 4651-4800, 4651-5127, 4652-5151, 4722-5146, 4736-5016, 4741-5213, 4746-4970, 4746-5207, 4746-5213, 4747-4965, 4750-5165, 4752-5213, 4754-4981, 4754-5215, 4756-5159, 4771-5217, 4779-5213, 4785-5076, 4787-5164, 4790-5032, 4811-5092, 4812-5211, 4818-5217, 4835-5100, 4844-5158, 4844-5165.

Table 4

Polynucleotide SEQ ID NO./ Incid ID/ Sequence Length	Sequence Fragments
20/500099CBI/ 1849	4854-5164, 4854-5217, 4863-5006, 4863-5197, 4863-5212, 4867-5163, 4867-5165, 4870-5057, 4881-5089, 4883-5138, 4885-5168, 4888-5149, 4907-5270, 4965-5219, 4966-5154, 4966-5217, 4991-5219, 4993-5164, 5050-5213, 5111-5217, 5283-5553, 5285-5525
20/500099CBI/ 1849	1-270, 1-375, 1-400, 1-475, 1-534, 1-536, 4-1847, 22-280, 28-293, 51-642, 123-754, 307-902, 411-1001, 434-969, 437-902, 447-1073, 450-1063, 485-1003, 487-1082, 498-988, 557-1131, 571-1069, 580-1191, 642-801, 729-1012, 810-1094,
21/668243CBI/ 1427	810-1098, 810-1102, 879-1039, 879-1391, 1100-1420, 1100-1628, 1100-1659, 1100-1647, 1100-1671, 1100-1775, 1101-1673, 1170-1782, 1241-1653,
21/668243CBI/ 1427	1251-1849, 1264-1849, 1271-1654, 1274-1704, 1345-1732, 1464-1757, 1475-1807, 1483-1738, 1615-1821,
22/2202389CBI/ 1014	1-575, 72-612, 260-466, 341-923, 341-944, 341-974, 341-976, 341-1072, 371-891, 373-934, 632-1085, 676-935, 676-1093, 759-1275, 856-1156, 856-1427
22/2202389CBI/ 1014	1-365, 1-510, 246-507, 246-739, 336-990, 509-1013, 550-1014, 556-1014, 574-1014, 593-1014, 599-1014, 605-1014, 734-1007, 769-899
23/7503597CBI/ 3695	1-642, 1-803, 1-820, 26-820, 54-890, 71-625, 197-821, 238-820, 618-1141, 618-1162, 620-806, 620-820, 849-1210, 922-1166, 922-1210, 945-1723, 946-1385, 946-1524, 946-1621, 946-1643, 946-1649, 946-1687, 946-1692, 970-1466, 1001-1770, 1006-1781, 1012-1354, 1015-1433, 1046-1792, 1046-3681, 1057-1524, 1060-1526, 1060-1846, 1205-1972, 1208-1731, 1211-1845, 1440-2053, 1920-2563, 1921-2563, 2015-2508, 2043-2278, 2043-2537, 2043-2599, 2043-2646, 2043-2657, 2043-2703, 2043-2759, 2043-2792, 2043-2872, 2044-2721, 2075-2563, 2236-2796, 2285-2563, 2473-3226, 2571-3367, 2572-3099, 2572-3135, 2572-3204, 2572-3223, 2621-3523, 2733-3345, 2749-2988, 2775-3626, 2823-3103, 2823-3360, 2824-3273, 2831-3445, 2853-3629, 2917-3607, 2920-3695, 2947-3695, 2951-3629, 2957-3291, 2964-3239, 2964-3638, 2967-3630, 2976-3504, 2981-3626, 2983-3692, 2984-3659, 2984-3636, 2995-3569, 2995-3693, 3010-3630, 3019-3630, 3029-3695, 3036-3425, 3044-3630, 3048-3521, 3049-3693, 3077-3695, 3109-3670, 3117-3695, 3126-3695, 3145-3676, 3180-3566, 3204-3695, 3205-3680, 3220-3695, 3222-3695, 3238-3681, 3255-3695, 3311-3695, 3314-3695, 3343-3695, 3365-3680, 3449-3695
24/7503603CBI/ 2403	1-212, 1-818, 1-829, 1-2397, 612-758, 612-815, 612-1152, 622-911, 737-1359, 746-1024, 755-930, 776-1303, 780-902, 797-1100, 897-1519, 925-1127, 925-1188, 925-1530, 990-1554, 1022-1376, 1066-1334, 1066-1712, 1075-1704, 1076-1669, 1127-1693,

WO 02/072794

PCT/US02/09052

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	1197-1820, 1197-1838, 1220-1800, 1236-1889, 1259-2127, 1264-2127, 1273-1740, 1296-1805, 1299-2127, 1314-1740, 1328-2127, 1346-1398, 1350-2127,
	1376-1428, 1377-2049, 1420-1934, 1470-2324, 1472-1968, 1473-2081, 1475-1679, 1478-2130, 1479-2075, 1513-2126, 1514-2323, 1531-2326, 1545-1939,
	1562-2093, 1579-2325, 1599-2214, 1600-2326, 1603-2127, 1643-2310, 1662-2326, 1689-2191, 1703-2357, 1709-2145, 1739-2343, 1845-2295, 1869-2317, 1895-2382, 1946-2273, 2040-2256, 2104-2403

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
13	3855123CB1	BRAHNON05
14	4547188CB1	COLXTDT01
15	3939883CB1	SKINBIT01
16	3163819CB1	TLYMTXT04
17	8518269CB1	TLYJTJXF01
18	1592646CB1	EOSIHET02
19	7500191CB1	BRAIFER05
20	7500099CB1	LUNGDIN02
21	7682434CB1	BRABDIK02
22	2202389CB1	SPLNFET02
23	7503597CB1	BRAHNON05
24	7503603CB1	COLXTDT01

WO 02/072794

PCT/US02/09052

Table 6

Library	Vectors	Library Description
BRADIK02	PSFORT1	This amplified and normalized library was constructed using pooled cDNA from three different donors. cDNA was generated using mRNA isolated from diseased vermis tissue removed from a 79-year-old Caucasian female (donor A) who died from pneumonia, an 83-year-old Caucasian male (donor B) who died from congestive heart failure, and an 87-year-old Caucasian female (donor C) who died from esophageal cancer. Pathology indicated severe Alzheimer's disease in donors A & B and moderate Alzheimer's disease in donor C. Patient history included glaucoma, pseudophakia, gastritis with gastrointestinal bleeding, peripheral vascular disease, chronic obstructive pulmonary disease, seizures, tobacco abuse in remission, and transient ischemic attacks in donor A; Parkinson's disease and atherosclerosis in donor B; hypertension, coronary artery disease, cerebral vascular accident, and hypothyroidism in donor C. Family history included Alzheimer's disease in the mother and sibling(s) of donor A. Independent clones from this amplified library were normalized in one round using conditions adapted Soares et al., PNAS (1994) 91:9228-9232 and
BRAHNON05	pNCY	Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used. This normalized hippocampus tissue library was constructed from 1.6 million independent clones from a hippocampus tissue library. Starting RNA was made from posterior hippocampus removed from a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. The cerebral hemisphere revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. There were small microscopic areas of cavitation with gliosis, scattered through the cerebral cortex. Patient history included cardiomyopathy, CHF, cardiomegaly, an enlarged spleen and liver. Patient medications included simethicone, Lasix, Digoxin, Colace, Zantac, captopril, and Vasotec. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAIFER05	pNCY	Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks gestation.
COLXTDT01	pNCY	Library was constructed using RNA isolated from colon tissue removed from the appendix of a 27-year-old Black female during myomectomy, dilation and curettage, right fibrial region biopsy, and incidental appendectomy. Pathology indicated an unremarkable appendix. Pathology for the associated tumor tissue indicated multiple (12) uterine leiomyomata. Patient history included premenopausal menorrhagia and sarcoedosis of the lung. Family history included acute myocardial infarction and atherosclerotic coronary artery disease.

WO 02/072794

PCT/US02/09052

Table 6

Library	Vector	Library Description
EOSINET02	PBLUESCRIPT	Library was constructed using RNA isolated from peripheral blood cells apheresed from a 48-year-old Caucasian male. Patient history included hypereosinophilia. The cell population was determined to be greater than 77% eosinophils by Wright's staining.
LUNGDI02	pINCY	This normalized lung tissue library was constructed from 7.6x10 ⁶ independent clones from a diseased lung tissue library. Starting RNA was made from RNA isolated from diseased lung tissue. Pathology indicated idiopathic pulmonary disease. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:5228-5232 and Ronaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
SKINBIT01	pINCY	Library was constructed using RNA isolated from diseased skin tissue of the left lower leg. Patient history included erythema nodosum of the left lower leg.
SPLNFET02	pINCY	Library was constructed using RNA isolated from spleen tissue removed from a Caucasian male fetus, who died at 23 weeks' gestation.
TLYJTXF01	PRARE	This 5' cap isolated full-length library was constructed using RNA isolated from a treated Jurkat cell line derived from the T cells of a male. The cells were treated with 5nM of PMA and 50ng/mL of Ionomycin for 1 hour. Patient history included acute T-cell leukemia.
TLYMTXT04	pINCY	Library was constructed using RNA isolated from CD4+ T cells obtained from a pool of donors. The cells were treated with CD3 and CD28 antibodies.

WO 02/072794

PCT/US02/09052

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABIPARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less; Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, fasta, fastx, fastx, and search.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:462-489.	ESTs: fasta E value= 1.06E-6; Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less; Full Length sequences: fastx score= 100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLocks, PRINTS, DOME, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less

WO 02/072794

PCT/US02/09052

Table 7

Program	Description	Reference	Parameter Threshold
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, SMART and TIGRFAM.	Kreigh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L., et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a NuShell, Cambridge Univ. Press, pp. 1-350.	PFAM: SMART or TIGRFAM hits: Probability value= 1.0E-3 or less; Signal peptide hits: Score= 0 or greater
ProfileScan	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; Baroch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality scores>GCG-specified "HGCI" value for that particular Prosite motif. Generally, scores= j,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 Phrap Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	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; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	

Table 7

Program	Description	Reference	Parameter Threshold
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sommarum, E.L. et al. (1998) Proc. Sixth Int'l. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Montis	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 02/072794

PCT/US02/09052

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 SEQ ID NO:1-12,
 - 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:1-6 and SEQ ID NO:8-12,
 - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and
 - d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.
2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.
3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method of 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

WO 02/072794

PCT/US02/09052

- polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
- b) recovering the polypeptide so expressed.
- 5 10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.
11. An isolated antibody which specifically binds to a polypeptide of claim 1.
- 10 12. 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:13-24,
- b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ
- 15 ID NO:13-18 and SEQ ID NO:20-24,
- c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 94% identical to the polynucleotide sequence of SEQ ID NO:19,
- d) a polynucleotide complementary to a polynucleotide of a),
- e) a polynucleotide complementary to a polynucleotide of b),
- 20 f) a polynucleotide complementary to a polynucleotide of c), and
- e) an RNA equivalent of a)-f).
13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.
- 25 14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:
- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and
- 30 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

WO 02/072794

PCT/US02/09052

present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

5 16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- 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
10 thereof, and, optionally, if present, the amount thereof.

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

15 18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

19. A method for treating a disease or condition associated with decreased expression of functional IGSEFP, comprising administering to a patient in need of such treatment the composition of
20 claim 17.

20. A method of 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.
25

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

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

WO 02/072794

PCT/US02/09052

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

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

5

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

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

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

- 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.

15

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

25

30

28. A method of 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:

WO 02/072794

PCT/US02/09052

- 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.
- 5
29. A method of assessing toxicity of a test compound, the 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 12 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 12 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 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.
- 10
- 15
- 20
30. A diagnostic test for a condition or disease associated with the expression of IGFSP in a biological sample, the method comprising:
- a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and
b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.
- 25
31. The antibody of claim 11, wherein the antibody is:
- a) a chimeric antibody,
b) a single chain antibody,
c) a Fab fragment,
d) a F(ab')₂ fragment, or
e) a humanized antibody.
- 30

WO 02/072794

PCT/US02/09052

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

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

5

34. A composition of claim 32, wherein the antibody is labeled.

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

10

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, or an immunogenic fragment thereof, under conditions to 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 specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

15

37. A polyclonal antibody produced by a method of claim 36.

20

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

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

25

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, 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

30

WO 02/072794

PCT/US02/09052

- e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.
- 5 40. A monoclonal antibody produced by a method of claim 39.
41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.
42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression
10 library.
43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.
- 15 44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12 in a sample, the method comprising:
- a) incubating the antibody of claim 11 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
20 polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12 in the sample.
45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12 from a sample, the method comprising:
- 25 a) incubating the antibody of claim 11 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 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.
- 30 46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

WO 02/072794

PCT/US02/09052

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:
- a) labeling the polynucleotides of the sample,
 - b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides
5 of the sample under conditions suitable for the formation of a hybridization complex, and
 - c) quantifying the expression of the polynucleotides in the sample.
48. An array comprising different nucleotide molecules affixed in distinct physical locations
10 on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.
49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is
15 completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.
50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.
- 20 51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.
52. An array of claim 48, which is a microarray.
- 25 53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.
54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.
30
55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains

WO 02/072794

PCT/US02/09052

nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.
57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.
58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.
59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.
60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.
61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.
62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.
63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.
64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.
65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.
66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.
67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.
68. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:13.
69. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:14.
70. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:15.

WO 02/072794

PCT/US02/09052

71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:16.
72. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:17.
- 5 73. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:18.
74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:19.
75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:20.
- 10 76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.
77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.
- 15 78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.
79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

WO 02/072794

PCT/US02/09052

<110> INCYTE GENOMICS, INC.
 YUE, Henry
 XU, Yuming
 THANGAVELU, Kavitha
 WARREN, Bridget A.
 TANG, Y. Tom
 DUGGAN, Brendan M.
 TRAN, Uyen K.
 BAUGHN, Mariah R.
 HONCHELL, Cynthia D.
 BURFORD, Neil
 FORSYTHE, Ian J.
 YANG, Junming
 MASON, Patricia M.

<120> IMMUNOGLOBULIN SUPERFAMILY PROTEINS

<130> PF-0925 PCT

<140> To Be Assigned
 <141> Herewith

<150> 60/275,249; 60/316,810; 60/323,977; 60/348,447;
 60/343,880
 <151> 2001-03-12; 2001-08-31; 2001-09-21; 2001-10-26;
 2001-11-02

<160> 24
 <170> PERL Program

<210> 1
 <211> 442
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3855123CD1

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

WO 02/072794

PCT/US02/09052

Glu Asn Val Leu Glu Ser Gly Thr Ser Gly Arg Tyr Thr Val Glu
 125 130 135
 Thr Ile Ser Thr Glu Glu Gly Val Ile Ser Thr Leu Thr Ile Ser
 140 145 150
 Asn Ile Val Arg Ala Asp Phe Gln Thr Ile Tyr Asn Cys Thr Ala
 155 160 165
 Trp Asn Ser Phe Gly Ser Asp Thr Glu Ile Ile Arg Leu Lys Glu
 170 175 180
 Gln Gly Ser Glu Met Lys Ser Gly Ala Gly Leu Glu Ala Glu Ser
 185 190 195
 Val Pro Met Ala Val Ile Ile Gly Val Ala Val Gly Ala Gly Val
 200 205 210
 Ala Phe Leu Val Leu Met Ala Thr Ile Val Ala Phe Cys Cys Ala
 215 220 225
 Arg Ser Gln Arg Asn Leu Lys Gly Val Val Ser Ala Lys Asn Asp
 230 235 240
 Ile Arg Val Glu Ile Val His Lys Glu Pro Ala Ser Gly Arg Glu
 245 250 255
 Gly Glu Glu His Ser Thr Ile Lys Gln Leu Met Met Asp Arg Gly
 260 265 270
 Glu Phe Gln Gln Asp Ser Val Leu Lys Gln Leu Glu Val Leu Lys
 275 280 285
 Glu Glu Glu Lys Glu Phe Gln Asn Leu Lys Asp Pro Thr Asn Gly
 290 295 300
 Tyr Tyr Ser Val Asn Thr Phe Lys Glu His His Ser Thr Pro Thr
 305 310 315
 Ile Ser Leu Ser Ser Cys Gln Pro Asp Leu Arg Pro Ala Gly Lys
 320 325 330
 Gln Arg Val Pro Thr Gly Met Ser Phe Thr Asn Ile Tyr Ser Thr
 335 340 345
 Leu Ser Gly Gln Gly Arg Leu Tyr Asp Tyr Gly Gln Arg Phe Val
 350 355 360
 Leu Gly Met Gly Ser Ser Ser Ile Glu Leu Cys Glu Arg Glu Phe
 365 370 375
 Gln Arg Gly Ser Leu Ser Asp Ser Ser Ser Phe Leu Asp Thr Gln
 380 385 390
 Cys Asp Ser Ser Val Ser Ser Ser Gly Lys Gln Asp Gly Tyr Val
 395 400 405
 Gln Phe Asp Lys Ala Ser Lys Ala Ser Ala Ser Ser Ser His His
 410 415 420
 Ser Gln Ser Ser Ser Gln Asn Ser Asp Pro Ser Arg Pro Leu Gln
 425 430 435
 Arg Arg Met Gln Thr His Val
 440

<210> 2
 <211> 577
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 4547188CD1

<400> 2
 Met Asp Gly Glu Ala Thr Val Lys Pro Gly Glu Gln Lys Glu Val

WO 02/072794

PCT/US02/09052

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

WO 02/072794

PCT/US02/09052

425 430 435
 Trp Thr Ser Met Glu Ala Ala Ser Gly Glu Gly Ser Ala Ala Gly
 440 445 450
 Asp Leu Asp Ala Ala Thr Gly Asp Arg Gly Pro Gln Ala Thr Leu
 455 460 465
 Ser Gln Thr Pro Ala Val Gly Pro Trp Gly Pro Pro Gly Lys Glu
 470 475 480
 Ser Ser Val Lys Arg Thr Phe Pro Glu Asp Glu Ser Ser Ser Arg
 485 490 495
 Thr Leu Ala Pro Val Ser Thr Met Leu Ala Leu Phe Met Leu Met
 500 505 510
 Ala Leu Val Leu Leu Gln Arg Lys Leu Trp Arg Arg Arg Thr Ser
 515 520 525
 Gln Glu Ala Glu Arg Val Thr Leu Ile Gln Met Thr His Phe Leu
 530 535 540
 Glu Val Asn Pro Gln Ala Asp Gln Leu Pro His Val Glu Arg Lys
 545 550 555
 Met Leu Gln Asp Asp Ser Leu Pro Ala Gly Ala Ser Leu Thr Ala
 560 565 570
 Pro Glu Arg Asn Pro Gly Pro
 575

<210> 3

<211> 385

<212> FRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No. 3939883CD1

<400> 3

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

WO 02/072794

PCT/US02/09052

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

<210> 4
 <211> 221
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3163819CD1

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

WO 02/072794

PCT/US02/09052

```

125          130          135
Val Leu Arg Lys Arg Arg Asp Ser Leu Ser Leu Ser Thr Gln Arg
140          145          150
Thr Gln Gly Pro Ala Glu Ser Ala Arg Asn Leu Glu Tyr Val Ser
155          160          165
Val Ser Pro Thr Asn Asn Thr Val Tyr Ala Ser Val Thr His Ser
170          175          180
Asn Arg Glu Thr Glu Ile Trp Thr Pro Arg Glu Asn Asp Thr Ile
185          190          195
Thr Ile Tyr Ser Thr Ile Asn His Ser Lys Glu Ser Lys Pro Thr
200          205          210
Phe Ser Arg Ala Thr Ala Leu Asp Asn Val Val
215          220

```

<210> 5

<211> 332

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 8518269CD1

<400> 5

```

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

```

WO 02/072794

PCT/US02/09052

Ile Ile Leu Leu Leu Leu Val Leu Arg Lys Arg Arg Asp Ser Leu
 245 250 255
 Ser Leu Ser Thr Gln Arg Thr Gln Gly Pro Ala Glu Ser Ala Arg
 260 265 270
 Asn Leu Glu Tyr Val Ser Val Ser Pro Thr Asn Asn Thr Val Tyr
 275 280 285
 Ala Ser Val Thr His Ser Asn Arg Glu Thr Glu Ile Trp Thr Pro
 290 295 300
 Arg Glu Asn Asp Thr Ile Thr Ile Tyr Ser Thr Ile Asn His Ser
 305 310 315
 Lys Glu Ser Lys Pro Thr Phe Ser Arg Ala Thr Ala Leu Asp Asn
 320 325 330
 Val Val

<210> 6

<211> 288

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No.: 1592646CD1

<400> 6

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

WO 02/072794

PCT/US02/09052

```

                230                235                240
Ser Trp Glu Asp Ser Gly Ser Ser Asp Tyr Thr Arg Gly Asn Leu
                245                250                255
Val Arg Leu Gly Leu Ala Gly Leu Val Leu Ile Ser Leu Gly Ala
                260                265                270
Leu Val Thr Phe Asp Trp Arg Ser Gln Asn Arg Ala Pro Ala Gly
                275                280                285
Ile Arg Pro

```

```

<210> 7
<211> 1450
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7500191CD1

```

```

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

```

WO 02/072794

PCT/US02/09052

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

WO 02/072794

PCT/US02/09052

Gln Leu Ser Gln Leu Ile Glu Gly Leu Asp Arg Gly Thr Glu Tyr
 695 700 705
 Asn Phe Arg Val Ala Ala Leu Thr Ile Asn Gly Thr Gly Pro Ala
 710 715 720
 Thr Asp Trp Leu Ser Ala Glu Thr Phe Glu Ser Asp Leu Asp Glu
 725 730 735
 Thr Arg Val Pro Glu Val Pro Ser Ser Leu His Val Arg Pro Leu
 740 745 750
 Val Thr Ser Ile Val Val Ser Trp Thr Pro Glu Asn Gln Asn
 755 760 765
 Ile Val Val Arg Gly Tyr Ala Ile Gly Tyr Gly Ile Gly Ser Pro
 770 775 780
 His Ala Gln Thr Ile Lys Val Asp Tyr Lys Gln Arg Tyr Tyr Thr
 785 790 795
 Ile Glu Asn Leu Asp Pro Ser Ser His Tyr Val Ile Thr Leu Lys
 800 805 810
 Ala Phe Asn Asn Val Gly Glu Gly Ile Pro Leu Tyr Glu Ser Ala
 815 820 825
 Val Thr Arg Pro His Thr Asp Thr Ser Glu Val Asp Leu Phe Val
 830 835 840
 Ile Asn Ala Pro Tyr Thr Pro Val Pro Asp Pro Thr Pro Met Met
 845 850 855
 Pro Pro Val Gly Val Gln Ala Ser Ile Leu Ser His Asp Thr Ile
 860 865 870
 Arg Ile Thr Trp Ala Asp Asn Ser Leu Pro Lys His Gln Lys Ile
 875 880 885
 Thr Asp Ser Arg Tyr Tyr Thr Val Arg Trp Lys Thr Asn Ile Pro
 890 895 900
 Ala Asn Thr Lys Tyr Lys Asn Ala Asn Ala Thr Thr Leu Ser Tyr
 905 910 915
 Leu Val Thr Gly Leu Lys Pro Asn Thr Leu Tyr Glu Phe Ser Val
 920 925 930
 Met Val Thr Lys Gly Arg Arg Ser Ser Thr Trp Ser Met Thr Ala
 935 940 945
 His Gly Thr Thr Phe Glu Leu Val Pro Thr Ser Pro Pro Lys Asp
 950 955 960
 Val Thr Val Val Ser Lys Glu Gly Lys Pro Lys Thr Ile Ile Val
 965 970 975
 Asn Trp Gln Pro Pro Ser Glu Ala Asn Gly Lys Ile Thr Gly Tyr
 980 985 990
 Ile Ile Tyr Tyr Ser Thr Asp Val Asn Ala Glu Ile His Asp Trp
 995 1000 1005
 Val Ile Glu Pro Val Val Gly Asn Arg Leu Thr His Gln Ile Gln
 1010 1015 1020
 Glu Leu Thr Leu Asp Thr Pro Tyr Tyr Phe Lys Ile Gln Ala Arg
 1025 1030 1035
 Asn Ser Lys Gly Met Gly Pro Met Ser Glu Ala Val Gln Phe Arg
 1040 1045 1050
 Thr Pro Lys Ala Ser Gly Ser Gly Gly Lys Gly Ser Arg Leu Pro
 1055 1060 1065
 Asp Leu Gly Ser Asp Tyr Lys Pro Pro Met Ser Gly Ser Asn Ser
 1070 1075 1080
 Pro His Gly Ser Pro Thr Ser Pro Leu Asp Ser Asn Met Leu Leu
 1085 1090 1095
 Val Ile Ile Val Ser Val Gly Val Ile Thr Ile Val Val Val Val
 1100 1105 1110

WO 02/072794

PCT/US02/09052

Ile Ile Ala Val Phe Cys Thr Arg Arg Thr Thr Ser His Gln Lys
 1115 1120 1125
 Lys Lys Arg Ala Ala Cys Lys Ser Val Asn Gly Ser His Lys Tyr
 1130 1135 1140
 Lys Gly Asn Ser Lys Asp Val Lys Pro Pro Asp Leu Trp Ile His
 1145 1150 1155
 His Glu Arg Leu Glu Leu Lys Pro Ile Asp Lys Ser Pro Asp Pro
 1160 1165 1170
 Asn Pro Ile Met Thr Asp Thr Pro Ile Pro Arg Asn Ser Gln Asp
 1175 1180 1185
 Ile Thr Pro Val Asp Asn Ser Met Asp Ser Asn Ile His Gln Arg
 1190 1195 1200
 Arg Asn Ser Tyr Arg Gly His Glu Ser Glu Asp Ser Met Ser Thr
 1205 1210 1215
 Leu Ala Gly Arg Arg Gly Met Arg Pro Lys Met Met Met Pro Phe
 1220 1225 1230
 Asp Ser Gln Pro Pro Gln Pro Val Ile Ser Ala His Pro Ile His
 1235 1240 1245
 Ser Leu Asp Asn Pro His His His Phe His Ser Ser Ser Leu Ala
 1250 1255 1260
 Ser Pro Ala Arg Ser His Leu Tyr His Pro Gly Ser Pro Trp Pro
 1265 1270 1275
 Ile Gly Thr Ser Met Ser Leu Ser Asp Arg Ala Asn Ser Thr Glu
 1280 1285 1290
 Ser Val Arg Asn Thr Pro Ser Thr Asp Thr Met Pro Ala Ser Ser
 1295 1300 1305
 Ser Gln Thr Cys Cys Thr Asp His Gln Asp Pro Glu Gly Ala Thr
 1310 1315 1320
 Ser Ser Ser Tyr Leu Ala Ser Ser Gln Glu Glu Asp Ser Gly Gln
 1325 1330 1335
 Ser Leu Pro Thr Ala His Val Arg Pro Ser His Pro Leu Lys Ser
 1340 1345 1350
 Phe Ala Val Pro Ala Ile Pro Pro Pro Gly Pro Pro Thr Tyr Asp
 1355 1360 1365
 Pro Ala Leu Pro Ser Thr Pro Leu Leu Ser Gln Gln Ala Leu Asn
 1370 1375 1380
 His His Ile His Ser Val Lys Thr Ala Ser Ile Gly Thr Leu Gly
 1385 1390 1395
 Arg Ser Arg Pro Pro Met Pro Val Val Val Pro Ser Ala Pro Glu
 1400 1405 1410
 Val Gln Glu Thr Thr Arg Met Leu Glu Asp Ser Glu Ser Ser Tyr
 1415 1420 1425
 Glu Pro Asp Glu Leu Thr Lys Glu Met Ala His Leu Glu Gly Leu
 1430 1435 1440
 Met Lys Asp Leu Asn Ala Ile Thr Thr Ala
 1445 1450

<210> 8
 <211> 551
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> misc_feature
 <223> Incyte ID No: 7500099CD1

WO 02/072794

PCT/US02/09052

```

<400> 8
Met Val Ala Pro Lys Ser His Thr Asp Asp Trp Ala Pro Gly Pro
1      5      10      15
Phe Ser Ser Lys Pro Gln Arg Ser Gln Leu Gln Ile Phe Ser Ser
20     25     30
Val Leu Gln Thr Ser Leu Leu Phe Leu Leu Met Gly Leu Arg Ala
35     40     45
Ser Gly Lys Asp Ser Ala Pro Thr Val Val Ser Gly Ile Leu Gly
50     55     60
Gly Ser Val Thr Leu Pro Leu Asn Ile Ser Val Asp Thr Glu Ile
65     70     75
Glu Asn Val Ile Trp Ile Gly Pro Lys Asn Ala Leu Ala Phe Ala
80     85     90
Arg Pro Lys Glu Asn Val Thr Ile Met Val Lys Ser Tyr Leu Gly
95     100    105
Arg Leu Asp Ile Thr Lys Trp Ser Tyr Ser Leu Cys Ile Ser Asn
110    115    120
Leu Thr Leu Asn Asp Ala Gly Ser Tyr Lys Ala Gln Ile Asn Gln
125    130    135
Arg Asn Phe Glu Val Thr Thr Glu Glu Glu Phe Thr Leu Phe Val
140    145    150
Tyr Glu Gln Leu Gln Glu Pro Gln Val Thr Met Lys Ser Val Lys
155    160    165
Val Ser Glu Asn Phe Ser Cys Asn Ile Thr Leu Met Cys Ser Val
170    175    180
Lys Gly Ala Glu Lys Ser Val Leu Tyr Ser Trp Thr Pro Arg Glu
185    190    195
Pro His Ala Ser Glu Ser Asn Gly Gly Ser Ile Leu Thr Val Ser
200    205    210
Arg Thr Pro Cys Asp Pro Asp Leu Pro Tyr Ile Cys Thr Ala Gln
215    220    225
Asn Pro Val Ser Gln Arg Ser Ser Leu Pro Val His Val Gly Gln
230    235    240
Phe Cys Thr Asp Pro Gly Ala Ser Arg Gly Gly Thr Thr Gly Glu
245    250    255
Thr Val Val Gly Val Leu Gly Glu Pro Val Thr Leu Pro Leu Ala
260    265    270
Leu Pro Ala Cys Arg Asp Thr Glu Lys Val Val Trp Leu Phe Asn
275    280    285
Thr Ser Ile Ile Ser Lys Glu Arg Glu Glu Ala Ala Thr Ala Asp
290    295    300
Pro Leu Ile Lys Ser Arg Asp Pro Tyr Lys Asn Arg Val Trp Val
305    310    315
Ser Ser Gln Asp Cys Ser Leu Lys Ile Ser Gln Leu Lys Ile Glu
320    325    330
Asp Ala Gly Pro Tyr His Ala Tyr Val Cys Ser Glu Ala Ser Ser
335    340    345
Val Thr Ser Met Thr His Val Thr Leu Leu Ile Tyr Arg Pro Glu
350    355    360
Arg Asn Thr Lys Leu Trp Ile Gly Leu Phe Leu Met Val Cys Leu
365    370    375
Leu Cys Val Gly Ile Phe Ser Trp Cys Ile Trp Lys Arg Lys Gly
380    385    390
Arg Cys Ser Val Pro Ala Phe Cys Ser Ser Gln Ala Glu Ala Pro
395    400    405
Ala Asp Thr Pro Gly Tyr Glu Lys Leu Asp Thr Pro Leu Arg Pro

```

WO 02/072794

PCT/US02/09052

```

410                               415                               420
Ala Arg Gln Gln Pro Thr Pro Thr Ser Asp Ser Ser Ser Asp Ser
425                               430                               435
Asn Leu Thr Thr Glu Glu Asp Glu Asp Arg Pro Glu Val His Lys
440                               445                               450
Pro Ile Ser Gly Arg Tyr Glu Val Phe Asp Gln Val Thr Gln Glu
455                               460                               465
Gly Ala Gly His Asp Pro Ala Pro Glu Gly Gln Ala Asp Tyr Asp
470                               475                               480
Pro Val Thr Pro Tyr Val Thr Glu Val Glu Ser Val Val Gly Glu
485                               490                               495
Asn Thr Met Tyr Ala Gln Val Phe Asn Leu Gln Gly Lys Thr Pro
500                               505                               510
Val Ser Gln Lys Glu Glu Ser Ser Ala Thr Ile Tyr Cys Ser Ile
515                               520                               525
Arg Lys Pro Gln Val Val Pro Pro Pro Gln Gln Asn Asp Leu Glu
530                               535                               540
Ile Pro Glu Ser Pro Thr Tyr Glu Asn Phe Thr
545                               550

```

<210> 9

<211> 336

<212> FRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7682434CD1

<400> 9

```

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

```

WO 02/072794

PCT/US02/09052

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

<210> 10

<211> 241

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2202389CD1

<400> 10

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

WO 02/072794

PCT/US02/09052

	185	190	195
Ser Asn Pro Cys Leu Glu Glu Asn Lys Pro Gly Ile Val Tyr Ala			
	200	205	210
Ser Leu Asn His Ser Val Ile Gly Leu Asn Ser Arg Leu Ala Arg			
	215	220	225
Asn Val Lys Glu Ala Pro Thr Glu Tyr Ala Ser Ile Cys Val Arg			
	230	235	240

Ser

<210> 11
 <211> 766
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7503597CD1

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

WO 02/072794

PCT/US02/09052

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

WO 02/072794

PCT/US02/09052

Ser Ile Glu Leu Cys Glu Arg Glu Phe Gln Arg Gly Ser Leu Ser
 695 700 705
 Asp Ser Ser Ser Phe Leu Asp Thr Gln Cys Asp Ser Ser Val Ser
 710 715 720
 Ser Ser Gly Lys Gln Asp Gly Tyr Val Gln Phe Asp Lys Ala Ser
 725 730 735
 Lys Ala Ser Ala Ser Ser Ser His His Ser Gln Ser Ser Ser Gln
 740 745 750
 Asn Ser Asp Pro Ser Arg Pro Leu Gln Arg Arg Met Gln Thr His
 755 760 765
 Val

<210> 12
 <211> 88
 <212> FRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7503603CD1

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

<210> 13
 <211> 2691
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3855123CB1

<400> 13
 ctccactggt caacccttct cgggtggagcc acagccaagt gctggaggac atacgtcgtc 60
 actttccact gctcttgcaa agccaaccc agctgtcacc cagtacaggt ggccaatgcg 120
 gggccagatc atcaaggagg catctggaga ggtgtcacc accacagtg actcacgta 180
 cttctcagag ccgctctcct gtgaggtag caacgcctgg gcagcaccaa cctcagccgc 240
 acggttgacg tctactttgg gcccgggatg accacagaac cccaatcctt gctcgtggat 300
 ctgggctctg atgcaatctt cagctgcgcc tggacgggca acccatcctt gaccatgctc 360
 tggatgaagc ggggctccgg agtggctcctg agcaatgaga agaccctgac cctcaaatcc 420
 gtgcgccagg agsacgagg caagtacgtg tgccgggctg tgggtccccg tgtgggagcc 480
 ggggagagag aggtgacctt gaccgtcaat ggacccccca tcatctccag caccagacc 540
 cagcacgccc tccacggcga gaaggccag atcaagtgct tcatccggag cagccgccc 600

WO 02/072794

PCT/US02/09052

```

ccggaccgca tgcctcgttc ctggaaggag aacgttctgg agteggggcac atcggggcgc 660
tatacggctg agaccatcag caaccgaggag ggcgtcatct ccacctgac catcagcaac 720
atcgtcgggg ccgacttcca gaccatctac aactgcacgg cctggaacag cttcggctcc 780
gacctcaaa tcattccggct caaggagcaa gtttcgaaa tgaagtggg agccgggttg 840
gaagcagagt ctgtgccgat ggcctctacc atbggggtag ccctaggaac tggtrggcc 900
ttctcgtctc ttatggcaac catcgtgggg ttctgctgtg cccttccca gagaatctc 960
aaaggtgttg tgcagccaa aaatgatata cyagtggaaa ttgtccaca ggaaccagcc 1020
tctggtcggg agggtaggaa gcactccacc atcaagcagc tgatgatga cgggggtgaa 1080
tlccagcaag actcagtcct gaacacagct gaggtctcca aagaagagga gaagagttt 1140
cagaacctga aggacccccc caatggtctc tacagctcca acacctcaa agagcaccac 1200
tcaaccccca ccattccctc ctccagctgc cagcccgacc tgcgtcctgc gggcaagcag 1260
cgtgtcccca caggcaatgt ctccaccaac atctacagca ccctgagcgg ccaggggcgc 1320
ctctacgact accggcagcg gtttgtgctg ggcattggga cctcgtccat cgagctttgt 1380
gacggggagt tccagagagg ctccctcagc gacagcagct cctccttga caagcagttg 1440
gacagcagcg tcaagcagcg cggcaagcag gatggctatg tgcagttga caaggccagc 1500
aaggcttctg ctctcctctc ccaccactcc cagtcctctg cccagaactc tgaccaccag 1560
cgaccctctc agcggcggat gcagactcac gtctaaagat cacacaccg ggtgggggac 1620
ggccagagga agaggtcagc gcacgttctg gttgtccagg gacgaggggt accttgccaga 1680
ggacaccaga attgcccaat tccaggacag cctccagcgg cctctgccac tgccttctt 1740
gaagctctg atcaagcaca aatctgggtc cccaggtgct gtgtgcaga ggtggggcgg 1800
tgggagagca gacagaggtc gggctgagt gcgctgtgct tagtctgga cacctctgc 1860
ccggccctt tctggagggc cctctacca cctgctctgc ccacaggac aagtggcage 1920
tataactctg ctctcaatgaa actgaggtcc actctctggt ctctctgtg gctctaccc 1980
tcaactgacca caagctctac ctaccctctg gctctgtctc caatacagcc ctggaggaaa 2040
gggatgagc tcttcccagc actgagctgc cccagaaac ccggctccc actgctgctc 2100
atagcccata cctcgaggcg tgacaagcca gaaatggcct tggctaaagg agcctctctc 2160
tcaccagctc ggcgggagc ccaccccaa tttgtttggt gttttgtgc cactactctg 2220
cagtlctctc ctgtgacttg atgcctctga actctcgggt gggaccggct ccctcagagc 2280
ctggttactc gggggagggg agggagggag gagcctgtgc tgacggagca cctcggcgg 2340
tgtcccctcc ctggctctgt tgaccaccgc ctcccacc ccctcctgct ttgtgtactc 2400
ctcccctccc cctcagcaca atcggagttc atataagaag tgcgggagct tctctgtcca 2460
gggttctctg aacacttatg gagaadcttc ttctctggaa gtgtgctgt tgaaggggct 2520
ggagggcagg tcttbaagat ggcgagactg ccctctcag ctgataaaca caagaacggc 2580
gatcctgtct tcaataggc tccacgagaa gagaggaagt atatctacac ctcaaccctc 2640
ctagtcaaca cctgaaataa atgttaggga cactacaaaa aaaaaaaaa a 2691

```

```

<210> 14
<211> 2518
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 4547188C81

```

```

<400> 14
ggaagatgat ggatcaatgt tttctttttt gaagctactg ttaccactcc tggaaaagt 60
cttcaggaaat aagtacagat aagaatgaca agggattagg actggcttcc tcttataaat 120
aataaaatcc aaagagaagt gacttgagtc tccaggttta aagaagagca actagaagtc 180
gtccaaacac ctgcattcca taaggagaag aaaagtccac ctggatttg tttctggact 240
gagatggatg gagagggcac agtgaagcct ggagaacaaa aggaagtgtt gaggagagga 300
agagaagtgg actactccag gctcattgct ggcactttac cacaaatcca cgtcaccagc 360
aggagggcag gatggaaaat gccctctctc ctacactgct gcctgctaca aggttctctc 420
ttccccttc cacaaaaaag accccatccg agatggctgt gggagggctc tctcccctcc 480
aggaccctc tccggccatc gggacactcc aggcctctct cgcctctctg ctgcccggag 540
gagagctcct ttgcagctcc aaattcattg aagggtccaa ggctggtgtc aggggagcct 600

```

WO 02/072794

PCT/US02/09052

```

ggaggagctg tcaaccatca gtgccattat gcccctcat ctgtcaacag gcaaccagag 660
aagtactggg gctgtctggg gcccocaaga tggatctgcc agaccattgt gteccaacac 720
cagatatactc accatcgtca tctgtaccgt gtggccctca cagactttcc acagagagggc 780
ttattttgtg tgaagctgtc ccaactgtcc cgggatgaca tggatgtcta cctctgccc 840
atgggaagtg aaaaacaacat gctgttttta agcatgaatc tgaccatctc tgcaggctcc 900
gccaagacc ccccaacagc cactccagct gctggggagc tcaccatgag atcctatgga 960
acagctcttc cagtggccaa cagatggacc ccaggaaaca cccagacctt aggcagaggg 1020
acagcatggg acacagttgc ttccactcca ggaaccagca agactacagc ttcagctgag 1080
gaaagacgaa ccccagggagc aaocagygca gcagctccag ggaacagcag ctgggagagag 1140
ggctctgtca aagcaccctgc tccgatccca gagagtcacc ctccaagag cagaagcatg 1200
tccaatacaa cagaaggtgtc ttgggagggc accaagagct cggtagacaaa cagggtcaga 1260
gccaagcaag acagagggga gatgacaact accaagggctg ataggccaag ggaggacata 1320
gagggggctc ggatagctct tgatgcagcc aaaaaggctc taggaacctt tggccaacca 1380
gctctggctc cagaaacttt ggccctgggaa atcctccacc aagcaacgcc agtttctaag 1440
caacaacttc agggttccat tggagaaaca actccagctg caggcatgtg gaccttggga 1500
actccagctg cagatgtgtg gatcttggga actccagctg cagatgtgtg gaccagcatg 1560
gagcgacatc ctggggaagg aagcgtctga ggggacctag atgctgccac tggagacaga 1620
ggcccccaag caacactgag ccagaccctg gcagtaggac cctggggacc ccttgccaag 1680
gagctctccg tgaagcgtac ttttccagaa gatgaaagca gctctggac cctggctcct 1740
gtctctacca tgcctggcct gtttatgctt atggctctgg ttctattgca aaggagctc 1800
tggagaaggc ggaaccttca ggagggagaa agggctcact taattcagat gacacattt 1860
ctggaagtga acccccagc agaccagctg ccccatgtg aaagaaagat gctccagat 1920
gactcttc ctgctggggc cagcctgact gcccagaga gaaatccagg accctgaggg 1980
acagagatg gaactgtca gttaccatg gagaagacc aagatcaag gccttcagga 2040
ccccagctc tttccatcat ccttctcca cctgtgggaa gagaagctga tgcagccgt 2100
gctccaccca tggaaagaa gctggctgtc ctggggcca agaaagctca gcaattacca 2160
cgtcccaagg tgacaagatg actcaaaagga gacttcaaga acagtgatg aaacactgga 2220
agaggtcacc taggaaaaag atgaaatttc catctcgaa tgtttgcaaa tgaagagggc 2280
ttccaactcg tgtgaaagt gacaaatccc ctatcaaac tccagccct tctgtggggc 2340
tcctttctg actactgtta gcaactcagc tccattcac atgtattata ttaagtgtg 2400
ccagcctgct cttctcaagt agattctaag ctctttaag gcagtaattg cattttatct 2460
gtctcatgat gcccacagag aacttcaac tcagtagacc ccaataatac ctgtgtgc 2518

```

<210> 15

<211> 1522

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3939883CB1

<400> 15

```

aaaccagtat tatgcaaac tcatccaac cctctgatt cottaacttg gctaagaaaa 60
agaggAagtt ctccagtta ctcaaccctg tggttctact atgctctctg accccgtctt 120
ggaactcaac tgggagaatg tggagccatt tgaacaggct cctctctctg agcatattt 180
cttctgtcac ttgtagaaaa gctgtattgg atgttgaggc aatgaaaaa aatgaattc 240
cttctccatg ttggactca aagactaagg tggttatgaa gggtaaaaat gtatctatgt 300
ttgttccca taagaacaaa tcaactcaga tcaactatc attgtttcga cgtaaagcac 360
acctgggaac ccaggatgga aaaggtgaac ctgcgattt taacctaac atcacagaag 420
ccatgaaac agcccccac aaatgcaag cccaagtac cagctgttca aatacagtc 480
gtgacttcag ctteccgatt gtcgaccggg tgaactcccc agtctgtaac attatggtca 540
tcaaacaga aacagaccga catataacat tacattgctt ctcagtaaat ggtctgctgc 600
ccatcaatta cactttcttt gaaaaccatg ttgccaatc accagctatt tccaagtatg 660
acagggagcc tgcctgaatt aacttaacca agaaatccc tgggagagag gaagagtata 720
ggtgtgagc taanaacaga ttgcctaact atgcaacata cagtcacct gtcaccatgc 780

```

WO 02/072794

PCT/US02/09052

```

cctcaacagg cggagacagc tgtcctttct gctggaagct actaactcca gggttattac 840
tgttgcctgt gctgataatc ctaattctgg ctttttgggt actgcccaca tacaaaaaca 900
gaaaagctat gagaataat gttcccagg accctggaga cacagccatg gaagtggaa 960
tctatgcaaa tatccttgaa aaacaagcaa agggagaaac tggccagaa gtggatcca 1020
ggccgtgtgt tccacagcc caagatgag ccaaacctc ccaggagcta cagtatgcca 1080
ccccgtgtt ccaggaggtg gcaccaagag agcaagaagc ctgtgattot tataaatctg 1140
gabatgtcta tctgaaact aaactctgaa atttacagaa acaactaca tctcaggatg 1200
gagtctcact ctgttgccca ggctggagtt cagtgggcgg atcttggctc actcaactc 1260
ccatctccc agttcaagcg attctcatgc ctgcacctc cagtagctg ggattcagg 1320
tgcctgctac cagcccagc taatttttgt atttttagta gagatgggtt tccactatgy 1380
tggccagctc ggtcttgaac tctgacctc agatgatcty cctgctcgg cctcccaag 1440
tgctggaact acaggcctga gccaccgtgc ccggccctga atcgcttag taagttaaag 1500
gtctccaaga ataaaaaaaa aa 1522

```

```

<210> 16
<211> 1084
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 3163819CB1

```

```

<400> 16
gaaaagcatg ttgtggctgt tccaatcgtc cctgtttctc ttctgctttg gcccaggaca 60
actgaggaac atacaagtta ccaatcacag tcagctattt cagaataga cctgtgact 120
ccatctcact tgcctctgtg aggatgcaga tgacaatgct tcattcagat gggaggcctt 180
ggaaacacaca ctttcaagtc agccaaactc cactgtctcc tgggacccca ggatttccag 240
tgaacaggaac tacacctgca tggccagaaa tgcgtcagat aatttatcct tctctgtctc 300
tgcctcagaag ctttgcgaag atgttaaaat tcaatataca gataccacaa tgattctgtt 360
tatggtttct gggatataca tagtctctgg tttcatcata ctgctgttac ttgltttgag 420
gaaaagaga gatccctat cttgtctac tcaggcaaca cagggccccg cagagctcgc 480
aaggaacctc ggtatgttt cagtgtctcc aacgaacac actggtgatg cttcagtcac 540
tcattcaaac agggaaacag aaatctggac acctagagaa aatgatacta tcacaattta 600
ctccacaatt aatcattcca aagagagtaa acccactttt tccagggcaa ctgcccctga 660
caatgtctgt taagtctctg aaaggcctca gaggaaatcg ggaatgacac gtctctgtat 720
cccatgagac agaacaaga acaggaagct tggttcctgt tgttctctgg aacagaattt 780
gaatatctag gataggatga tcacctccag tctctgggac ttaaacctgc ctacctgagt 840
caaacacctc aggtatacat catttccagc atgtgtgtca aataatattt tccaatccac 900
ttcagcccaa aacatgctaa agataacaca ccagcacatt gactctctct ttgataacta 960
agcaaatgga atlatggttg acagagagtt batgatccag aagacaacca cttctctcct 1020
tttagaaagc agcaggattg acttattgag aaataatgca gtgtgttggg tacatgtgta 1084
gtct 1084

```

```

<210> 17
<211> 1463
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 8518269CB1

```

```

<400> 17
caaaaacatt gactgcctca aggtctcaag caccagtctt caccgaggaa agcatgttgt 60
ggctgttcca atcgtcctg tttgtctctt gctttggccc agggaatgta gttccaaaa 120

```

WO 02/072794

PCT/US02/09052

```

goagctaac cccattgat gtgaacggga ttetggggga gtcagtaact ettcccctgg 180
agtttccctgc aggagagaag gtcaacttca tcacttggct tttoaatgaa acatctcttg 240
ccttcatagt accccatgaa accaaagtc cagaatcca cgtgactaat ccgaacacgg 300
gaaagcgcact gaacttccac cagctctact ccttgcaact cagcaacctg aagatggaag 360
acacaggctc ttaacagacc cagatataca caaagacctc tgcnaagctg tccagttaaa 420
ctctgaggat attaagacaa ctgaggaaaca tacaagttac caatcacagt cagctatttc 480
agaatataac ctgtgagctc catctgactt gctctgtgga ggtgcugat gacaatgctc 540
cattcagatg gggggccttg gaaacacac ttccaagtca gccaaacctc actgtctcct 600
gggaccocag gatttccagt gaacaggact acacctgcat agcagagaat gctgtcagta 660
attatccctt ctctctctct gccacagaag tttgccaaga tgtttaaatt caatatacag 720
ataccaaaat gattctgttt atggtttctg ggatagcat agtctccgtt ttcatacacc 780
tgctgttaact tgttttgagg aaaagaagag attccctatc tttgtctact cagcaaacac 840
agggccccgc agagtccgca aggaacctag agtatgtttc agtctctcca acgaacaaca 900
ctgtgtatgc ttcagtcaact cttccaanaa gggaaacaga aatctggaca cctagagaaa 960
atgatactat cacaatttac tcccaatta atcattcca agagagttaa cccacttttt 1020
ccagggcaac tgccttgcac aatgtcgtgt aagtgtgca aaggcctcag aggaattcgg 1080
gaatgacaag tttctgtatc ccatgagaca gaacaaagaa caggaagctt ggttctctgt 1140
gttcttgcca acagaatttg aatctctagc ataggatgat caccctccag cctctggact 1200
taaacctgcc taccctgagtc aaacacctaa ggataacatc atttccagca tggttctcaa 1260
ataatatttt ccaatccact tcaaggcaaa acatgtctaa gataacacac cagcaacttg 1320
actctctctt tgataactaa gcaaatggaa ttatggttga cagagagttt atgatccaga 1380
agacaaccac ttctctcctt ttgaaaagca gcaggattga cttattgaga aataatgcag 1440
tgtgttgggt acatgtgtag tct 1463

```

<210> 18

<211> 1557

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1592646CB1

<400> 18

```

agcggggc ac tgcgcagaa caaagatgga gccgtggagt gccatagggc tatgacacag 60
tccccacag gccccacct cgtactctgc ttccgtaaaat gaggatctgg gtctgggttt 120
ctgatgttgc ctcatttctt gggaggggag agggctgcac caagccctgg ctccagctct 180
agcgggtatc tgcaccaact ggcctgggtg ctgactctcc agctgctgac cctctggcct 240
ctgtgtcaaa cagacatcac tccgtctgtc cccccagctt cataccacc taagccatgg 300
ctgggagctc agcgggtcac agttgtgacc cctgggttca acgtgacctt gagatgccg 360
gcaccccaac ccgcttggag atttggactt ttcaagcctg gagagatcgc tcccctctc 420
ttccgggatg tglcctccga gclggcagaa ttctttctgg aggaggtgac tccagcccaa 480
gggggaagt acccctgctg ctaccgaagg ccagactggg ggcgggtgt ctggteccag 540
cccagcagtg tccctggagct gctgggtgaca gaggagctgc cgcggcctgc gctgggtggc 600
ctgcccgggc cgttgggtgg tccctggccc aacgtgagcc tgcctgccc gggccgctg 660
cggaaactga gcttctgtct gtaccgcyag ggcctggcgg ccccgtgca gtaccgccac 720
tcgcgcagc cctgggcgca ctaccgctg ctggggcggc gcgccccgg caccacacgc 780
tgtactatc acagcgcctc cgcgcctcac gtgctgtcgc agcgcagcga ggtgctggtc 840
atcagctggg aagactctgg ctccctccac tacaccggg ggaacctagt ccgctgggg 900
ctggccgggc tggctctcat ctccctgggc gcyctggta cttttgactg gcgcagtcag 960
aaaccgctc ctgctggtat ccgcccctga gcccccagg cactgcagcc cgaactctc 1020
aacctgagtg cggggaagc tgggacctg ggcctggact tcttctctg cagccccaca 1080
gtctctctg ctgagctccg cgggaacgct cttagacccc gctgtctct gtctgtgagc 1140
ttcttccag cctttccca aggagtagct gaaaggaaga cgcgattagt ggttaagact 1200
tccaagccag aagacagagg gttcgeatcc cagcactgcc gtactactac tgtatagta 1260
gcagctacag aaaggtagta gtgagcgtg aagccagctg gacttctctg gttgaatgg 1320

```

WO 02/072794

PCT/US02/09052

```

gacctggaga acctttctgt cttacaagag gattgtmaaa tggaccaatc agcactctgt 1380
aagatggacc aatcagcgcct ctgtaaaatg gaccaatcag caggacatgg gcggggacaa 1440
taagggaata aaagctggcg agcggggcac cccaccagag tctgcttcca cgtgtggga 1500
gctttgttct cttgctctac acaataaatc ttgctgctgc taaaaaaaaa aaaaagg 1557

```

```

<210> 19
<211> 5553
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7500191CB1

```

```

<400> 19
tgcggcgcgc gtagccgagc ttgcagcgag ggaccggctg aggcgcgagg gagggaaagga 60
ggcaaggcct ccgcccgcct gtgcgccgcg ctgcgcctca ctctcgggga agagatggcg 120
gcggagccgg gaccggcgct actcctcagc acccccctct tctggctcta ctgctctgtg 180
ctgctcgggc gccggggccc gggcggccgc gccgccagga gcggctccgc gccggcagtc 240
ccaggagcca gcaatcgaac gtctactcca ttttatcttc tggggagacc ggtggataca 300
ctctcagtta gaggctcttc tgttatatta aactgttccg catattctga cctctctcca 360
aaaattgaat ggaaaaaaga tggaaacttt ttaaacttag tatcagatga tcgacgcagc 420
ctctcccggc atggatcttt attlatcagc aatgtgtgca attcacaaca caataaaact 480
gatgaaggtt attatcagtg tggggcactt gttgagatgc ttggaactat latcagtaga 540
acagcgaagc tcatagtagc aggtcttcca agatttaca gccaccaga acctctcca 600
gtttatgctg ggaacaatgc aattctgaat tgtgaagtta atcgagattt ggtccacttt 660
gtgagtgagg aacagaacag acaaccctct cttctggatg atagagttat caaacttcca 720
agtgaatgac tggttatcag caatgcaact gaaggagatg gcggccttta tcgctcgtta 780
gtgaaaatg gtggccacc aaagtatagt gatgaagttg aattgaaggt tottcagat 840
cctgaggtga taccagactt ggtatctttg aaacagcctt ctcccttagt cagatcatt 900
ggtcaggatg tagtgttgc atgtgttgc tccaggacttc ctactccaac catlaaatgg 960
atgaaaatg agggagcact tgaccacgaa agctctgaaa gatttggatt cctggcaggt 1020
ggtagcctgg agatcagtga tgttactgag gatgatgctg ggacttattt tcttatagct 1080
gataatgaaa agcctactaa tatatagct cacgaatca tggatattgt atttgaatgt 1140
tctctgagc agcctactaa tatatagct cacgaatca tggatattgt atttgaatgt 1200
gaagtgaact gaaaaacca tccaactgtg aagtgggtca aaaaagggga tatggttatc 1260
ccaagtgat attttaagat tggtaaggaa cataatcttc aagttttggg tctggtgaaa 1320
tcagatgaag ggttctatca gtgcattgct gaaaatgatg ttggaaatgc acaagctgga 1380
gcccaactga taactcctga acatgcacca gccacaacgg gaccactgac ttcagctcct 1440
cgggalgtcg tggcctccct ggtctctacc cgttctacca aattgacgtg ggggacactt 1500
gcatcagatc ctccagcaga caaccttacc tactctgtgt tctcacacca ggaaggatt 1560
gctagggaac gtgttgagaa taccagtcac ccaggagaga tgcagtaac cattcaaac 1620
ctaagtccag cgaaccgtga catctttaga gttatggctc aaaaatagca tggctcagga 1680
gagagttcag ctccactgcy agtgaanaa caacctgagg ttcagctccc tggcccagca 1740
cctaaccttc gtccatagc agcttcyctt acctccatca ctgttacgtg ggaacacca 1800
gtgtctgcca atggggaat tcagaattat aaattgtact acatggaaaa gggactgat 1860
aaagaacagg atgttgatgt tccaagtac tcttacacca ttaattgggtt gaaaaaat 1920
acagagtata gtttccagat ggtggcctac aataaacatg gtcctggagt tccacacca 1980
gatgttctgt ttgcacact gtccagatgt cccagtcgct ctctcagaa tctgtctctg 2040
gaagtgaaga attcaaaagc tatltagatt cactggcagc cacctgctcc agccacaaa 2100
aatggcagaa ttactggcta caagattcgc taccgaagg cctccgaaa gactgatgct 2160
actgagacct tggtaagcgg gacacagctg tctcagctga ttgaaggtct tgatcgggg 2220
actgagtata atttccaggt gctgctctca acaatcaatg gtacaggccc ggcaactgac 2280
tggctgtctg ctgaaacttt tgaagtgcac ctgagatgaa ctgctgttcc tgaagtgcct 2340
agctctcttc acgtacgccc gctcgttact agcatcgtag tgagctggac tctccagag 2400
aatcagaaca ttgtgtcag auyttacgcc attgttatg gcattggcag cctccatgcc 2460

```

WO 02/07294

PCT/US02/09052

```

cagaccatca aagtggacta taacacgccc tattacacca ttgaaaatct ggatcccaga 2520
tctcactatg tgattaccct gaaagcattt aataacgtgg gtgaaggcat ccccctgtat 2580
gagagtgctg tgaccaggcc tcacacagac actctcgaag ttgatttatt tgttattaat 2640
gtccatatac etccagtgcc agatcccact cccatgatgc caccagltgg agtccaggct 2700
tccattctga gtcctgacac catcaggatt accgtggccag acaactcgtt gcccaagcac 2760
cagaagatta cacactcccc atactacacc gtccggtgga aaaccaacat cccagcaaac 2820
accaagtaca agaattgaaa tgcacaccact ttgagttatt tggtagctgg tttaaagccg 2880
aatacactct atgaattctc ttgtatgggt accaaaggto gaagatcaag tacatggagt 2940
atgacagccc atgggaccac ctttgaatta gtcccgactt ctccaccocaa gtagtggact 3000
gttgtgagta aaggggggaa accataagacc ataattgtga attggcgacc tccctctgaa 3060
gccaatggca aaattacagg ttacatcata tattacagta cagatgfgaa tgcagagata 3120
catgactggg ttattgagcc ttgttgggga aacagactga ctccaccagat acaagagtta 3180
actcttgaca caccatacta ctcaaaaatc caggcacgga actcaaaagg catgggacc 3240
atgtctgaag ctgtccaatt cagaacacct aaagcctcag ggtctggagg gaagggaagc 3300
cggctgcccag acctaggatc cgaactacaaa cctccaatga gcggcagtaa cagccctcat 3360
gggagcccaca cctctcctct ggacagtaat atgctgtggt tcataattgt tctctgttggc 3420
gtcacaacca tctgtgggtt ttgtattatc gctgtctttt gtaccctctg taccacctct 3480
caccagaaaa agaaaacgagc tgcctgcaaa tcagtgaatg gctctcmeta taccaaaagg 3540
aattccaag atgtgaaacc tccagatctc tggatccatc atgagagact ggagctgaaa 3600
cccattgata agtctccaga ccccaacccc atcatgactg atactccaat tectcgcaac 3660
tctcaagata tcaacccagc tgacaactcc atggcacgca atatccatca aaggcgaat 3720
tcatacagag ggcataagtc agaggacagc atgtctacac tggctggaaq gggaggaatg 3780
agacaaaana tgaatgatgc ctttgactcc cagccacccc agcctgtgat tagtggccat 3840
cccatacatt cctctgataa cctccaccat catttccact ccagcagcct cgtctctcca 3900
gctcgcagtc atctctacca cccgggcagc ccatggccca ttggcacatc catgtccctt 3960
tcagacaggg ccaattccac agaattccgtt cgaatacccc ccagcactga caccatgcca 4020
gectctctgt ctcaaacatg ctgcaactgat ccccaggacc ctgaaggtgc taccagctcc 4080
tcttacttgg ccagctccca aggaggaagt tccaggccaga gtcttcccac tggccatggt 4140
cgcccttccc acccattgaa gagcttctcc gtgccagcaa tcccgcctcc aggcactccc 4200
accatgatc ctgcatgtcc aagcacacca ttactgtccc agcaagctct gaacctcac 4260
attcaactcag tgaagacagc ctccatcggg actctaggaa gtagccggcc tctatgcca 4320
gtggttgctc ccagtgcccc tgaagtgccg gagaccacaa gnatgttga agactccgag 4380
agtagctatg aaccagatga gctgaccaaa gsgatggccc acctggaag actaatgag 4440
gacctaaacg ctatcacaaac agcatgacga ccttcaccag gaactgactt caaacctgag 4500
tctggaagtc ttggaactta acccttgaaa acaaggaatt gtacagagta cgagaggaca 4560
gcaactgaga acacagaatg agccagcaga ctggccagcc cctctgtgta gggctggctc 4620
caggcatggc cactgcctt cccctggctc gctgggaaga agcctgtgtc gaggcagctt 4680
ccctllgect gctgatattc tgcaaggactg ggcaccatgg gccaaaattt tgtgtccagg 4740
gaagaggcga gaagtgcacc ctgcatttca ctttgtgttc aggcctgttc tttgtctgt 4800
gactgcatca cctttatgga gtgtagacat tggcatttat gtacaattt atttgtgtt 4860
tattttattt tacctcaaaa acaaaaaaac ccatccaaaa ccaaggaagt ccttgggtt 4920
ctccacaagt ggttgacatt tgaactgctg ttccaattat gtatggaaag tctttgacag 4980
tgtggctcgt tctctggggt ggctgttttt ttggtttcat ttttattttt taattctgag 5040
tcaatgcatc ctctaccagc tgttaatcca tcaactctgag ggggagggaaa tgttgcattg 5100
ctgtttgtaa ccttttttta ttattttttt attataatta ttaaaggcct gactcttccc 5160
tctcatcact gtgagattac agatctattt gaattgaatg aaatgtaaca ttgaaaagac 5220
ttgtttgtg ctctctgtgc agttctagta ttggggcggg tggggggctg ggggttgata 5280
atagaaaag gaggggctgc tgaggtctct tgaatgttct tgcactgta cttctctcca 5340
gaagcctgca gagaatggaa gcatctctct tattgtcctt tccctggcat tccatctta 5400
ttgtcaactc gttgcaactg gagtttgatt tggatctggt tttaaaatc tctctgcaa 5460
tagatgggtt tgaggattta cggccctgca tgtctgtgct atagcctggt aagaatgtcc 5520
atgtgagga gccacatggt gtattttetaa ctg 5553

```

<210> 20
 <211> 1849
 <212> DNA

WO 02/072794

PCT/US02/09052

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7500099CB1

<400> 20

```

aatagatcat catgggtggca ccaaagagtc acacagatga ctgggctcct gggccttct 60
ccagtaagcc acagaggagtc cagctgcaaa tattctcttc tgttctacag acctctctcc 120
tcttctgct catgggacta agagcctctg gaaaggactc agccccaaac gtgggtctag 180
ggatcctagg ggggtccctg actctcccc taacatctc agtagacaca gagattgaga 240
acgtcatctg gattgtctcc aaaaatgctc ttgctttcgc acgtcccaaa gaaaatgtaa 300
ccattatggt caaaagctac ctgggctgac tagacatcac caagtggagt tactccctgt 360
gcatcagcaa tctgactctg aatgatgcag gatectaca agcccagata aaccaaggga 420
attttgaagt caccaactgag gaggaattca cctgttctg ctatgagcag ctgcaaggagc 480
cccaagtcaac catgaagtct gtaaagggtg ctgagaactt ctctgtaac atcaactbaa 540
tgtctctcgt gaagggggca gagaaaagtg ttctgtacag ctggacccca agggaacccc 600
atgcttctga gtccaatgaa ggcctccattc ttccgtcttc ccgaacacca tgtgaccacg 660
acctgccata catctgcaca gcccaagaacc ccgtcagcca gagaagctcc ctccctgtcc 720
atgttgggca gttctgtaca gatccaggag cctccagagg aggaacaacg ggggagactg 780
tgttaggggt cctgggagag ccagtcaccc tgcccactgc actcccagcc tgccgggaca 840
cagagaaggt tctctggttg ttaaacacat ccatcattag caaagagagg gaagaagcag 900
caagggcaga tccactcatt aaatccaggg atccttaca gaacagggtg tgggtctoca 960
gccaggaactg ctccctgaag atcagccagg tgaagataga ggagcccgcc ccttaccatg 1020
ctcactgttg ctaagaggcc tccagcgtca ccagcatgac acatgtcacc ctgctcatct 1080
accgaacctga gagaaacaca aagctttgga ttgggtgttt cctgatggtt tgcctctgt 1140
ggcttgggat cttaagctgg tgcatttggg agcgaaaaag acggtgttca gtcccagctc 1200
tctgttccag ccaagctgag gcccccagg atacaccagg atatgagaag ctggacaccc 1260
ccctcagccc tgcccaggca cagcctcac ccacctcaga cagcagctct gacagaacc 1320
tcacaactga ggaagatgag gacagccctg aggtgcaca gcccatcagt ggaagatag 1380
aggatattga ccagtcact caggagggcg ctgacatga cccagcccct gagggccaag 1440
cagactatga tccctcact ccatatgta cgaagtga gtctgtgtt ggagagaaca 1500
ccatgtatgc acaagtgtc aacttacag gaaagacccc agtttctcag aaggaagaga 1560
gtcagccac aatctactgc tccatcagya aacctcaggt ggtgccaca ccacaacaga 1620
atgatcttga gattcctgaa agtccctacct atgaaaaatt cacctgaaag gaaaagcagc 1680
tgtctctctc ctctgggac cgtgggggtg gaaagtcaag tggacctcat ggggctggy 1740
gtctgcagac agaagcact cagaatttcc ttcagtgctc cagagatgcc tggatgtggc 1800
ccctccccct ccttctcacc cttaaggact cccaaaccca ttaatagtt 1849

```

<210> 21

<211> 1427

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7682434CB1

<400> 21

```

cgccgctctc gccgcgatgc cccccctgc gccggggccc cggtccggc ttctgcgcg 60
cgccgctctc gccgcgatgc cccccctgc gccggggccc cggtccggc ttctgcgcg 120
caactctctc gccgacaact acacagtgtg tgaagggtgac aacgcaacc tcagctgctt 180
catcgacgag cagctgaccc cgtggccctg gctgaaccgc tccacaatcc tgtatgccg 240
caatgacgag tgaaccagc acccgccggt gggctgctc atcaacacc ccgagaggtt 300
ctccatctc ataccgagg tgggctctg cgagcagggc ctctacacc gctccttca 360
gaccgccac cagccgtaca ccaactcaggt ctacctcatt gtcccagtc ctgcccagat 420

```

WO 02/072794

PCT/US02/09052

```

tgtgaacatc tcttgcctg tgacggtgaa tgaggggggc aatgtgaacc tctttgctt 480
ggcctgtggg cggccagagc ccacggtcac ctggagacag ctccagagcg gcttaccctc 540
ggaggggagag atcctggaga tctctgacat ccagcggggc caggccgggg agtatgagtg 600
cgtgaectcac aacgggggta actcggcgcc cgaacgcgcg cgtgtgtgtg teacagtcaa 660
ctatcctccg accatcacgg acgtgaccag cgcgcgcacc gccttgggcc gggccgccc 720
cctgcgctgc gaagccatgg cgttcccccc cgcggatctc cagtggtaca aggatgacag 780
actgtctgagc agcggcacgg ccgaaggcct gaagggtgag acggagcgca cccgctcgat 840
gcttctcttt gccaacgtga gcgcggggca ttacgggaac tatacgtgtc ggcgcccaa 900
ccgactggga gcgtccagcg cctccatgcg gctcctgcgc ccaggatccc tggagaactc 960
agccccgagg cccccaggc tcttggcctc cctctccgcg ctgggtggc tgtggtggag 1020
aatgtaggcg caaccagtg gagctcaact cccctcgag ggggctcag gccaaagtg 1080
agagaacagg gggagcaaga gccgtgggtc tctggggggc agaagagctc tggccacca 1140
aggaagaaga gagaggaga gaggaggagg cagaggaaga aagatcttca gagaacctc 1200
cactgtgagg gataacgcaa aattatgcat cttctacag ccattctcgc caccgttca 1260
cgtttccgat tgtgaccac tcccgcacc ccataccctc cctctctagc tcaggctgtc 1320
aactggcttg tgtgggtgtg ggtgtgtgag tgtgagcctg catgcatgtg taggtgtctg 1380
tgtctctgtt tgtgtgtgtg tgggggggtg ggtgggggga agggact 1427

```

<210> 22
 <211> 1014
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2202389CB1

```

<400> 22
cacagatgcc actggggtag gtaaacctgac ccaactctgc agcactcaga agacgaagca 60
aagccttcta ctgagcagt ttttccatca ctgatatgtg caggaaatga agacattgcc 120
tgccatgctt ggaactggga aattattttg ggtctcttct ttaatcccat atctggaact 180
ctggaacatc catgggaaag aatcatgtga tgtacagctt tatataaaga gacaactctga 240
acactccatc ttagcagtag atccctttag actagaaatg cctgtgaaat actgtgctaa 300
caggcctcat gtgacttggg gcaagctcaa tggacaacaa tggtaaaac ttgaagatag 360
acaacaagt tggaaggaa agaaagaacat ttcatcttct atctacatt ttgaaccagt 420
gcttctaat gacaatgggt cataccgctg ttctgcaaat tttcagctca atctcattga 480
aagccactca acaactcttt atgtgacagg aaagcaaaat gaactctctg acacagcagg 540
aagggaaatt aacctggtg atgtcacct taagagttag caaacagaag caagcaccag 600
gcaaaatccc caagtactgc tatcagaaac tggaaattat gataatgacc ctgacctttg 660
ttcaggatg caggaaaggt ctgaagttaa ttcaatcca tgccctggaag aaaaacaacc 720
agccattggt tatgctccc tgaaccatc tgtcaatgga ctgnaactca gactggcaag 780
aaatgtaaaa gaagcaccaa cagaatatgc atccatagt gtgaggagtt aagctctgtt 840
ctgactccaa caggaccac tgaatgatca gcattgtgac atcattgtct gggctcaaca 900
ggaatcmeta taatattct caatttgaga attttactt tagaaatgtt catgttagtg 960
cttgggtctt aagggtccat aggataaatg attaaaattt ctctcagaaa ctta 1014

```

<210> 23
 <211> 3695
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7503597CB1

<400> 23

WO 02/072794

PCT/US02/09052

```

ccgcctgag gaagccgtgt gccctgggatg ccaagagcca gagaatggat cttctccgag 60
tgggacaatt cctgacaatc ccgcttccc gggccgcta agaaccggca gtttgtgtcg 120
gctggctgca gataccocaga gccacaaga gaccgaagcc acccggaggg acccaccgac 180
ggacagatgg taggcgcgaa ccgcsagaga ccggccggagg ctgagcaccg agagccocca 240
aggaagagaa actaacccac gccaaagtac ccggccggct tctcttctgt gcactaagga 300
atgaaacctt tocagctoga totgtctctc gtctgtctct tctcttctag tcaagagctg 360
ggctccaga agagagatc ctgtctgtg ctgggtaca tggccaagga caagtctcgg 420
agaatgaatg aaggccaagt ctattctctc agccagcagc cccagagcca gttgtgtgtg 480
tcgggacagc cagtgaagct actttgtgct atccccgaat acgattggct cgttctgtgg 540
atcaagagagc gcttggctct ggtgtgggg agggacctct caagtaccac acagtacctg 600
gtggtaggga accacctgtc aggggagcac cacctgaaga tccctgagggc agagctgcaa 660
gacgatgcgg tgtaccagt cccagccatc caggccocca tccctctccg ccccgaccgc 720
ctcacagtcc tgggtccgct tgatgacccc gtcactctgg gggccctctg gatcagctcg 780
cgtgcccggg accctctcaa cctcaccctc cagcagacaa atgccaagcc tgcagctctc 840
atcatctggt tgcgaaaggg agaggtctac aatggggcca cctactccaa gaccctgctt 900
cgggacgcca agcgggagag catctctcag accctcttca tctccccctg tgcagctggg 960
aatggccaga gcatcgttgt tcgtgccacc aacaaagcca tccccggagg aaaggagagc 1020
tcggtaacca ttgacatcca gccacctcca ctggtcaacc tctcgttgga gccacagcca 1080
gtgctggagg acaacgtcgt cactttccac tgccttgcaa aggcccaacc agctgtacc 1140
cagtacaggt gggccaagcg gggccagatc atcaaggagg catctggaga ggtgtacagg 1200
accacagtgg actaacagta cttctcagag ccctctctct gtgaggtgac caaccctctg 1260
ggcagcacca accctcagcc caccgttgac gttactcttg gcccgggat gaccacagaa 1320
ccccatctct tgcctgtgga tctgggtctc gatgccatct tcaagctgag cttgaccggc 1380
aaacctccc tgaccatcgt ctggatgaag cggggctccg gagtgtctct gagcaatgag 1440
aagacctga cctcaaatc cgtgcgccc gaggagcggg ccaagtactg gtgcccggct 1500
gtgtgcccc gttgtggagc cggggagaga gaggtagccc tgacctcaa tggcccccc 1560
tcatctcca gccaccagac ccagcagccc ctccaagggc agaagggcca gatcaagtgc 1620
tcatctcggg gccaccgccc ccggcagccc atcctctggt cctggaagga gaactgtctg 1680
gagtccggca catcggggcg ctatacgtg gagacctca gccaccagga gggctctcag 1740
tccacctga ccatcagcaa catctgccc gccactctcc agacctcta caactgacg 1800
ccctggaaa gcttgggctc cgacctgag atcatccggc tcaaggagca agagtctgtg 1860
ccgatggcg kcatcaktgg ggtggccgta ggagctgggt tggcttctc cgtctctatg 1920
gcaacctag tggcgtttgt ctgtgcccgt tcccagagaa atctcaagg tgttgtgtca 1980
gccaaaaatg ataccagat ggaatctgt cacaaggaac cagcctctgg tcgggagggg 2040
gaggagcaat ccaatcaaa ccagctgatg atggaccggg gtaattcca gccagactca 2100
gtcctgaaac agctggaggt cctcaagaa gaggagaaag agtttcagaa cctgaaggac 2160
cccaccaatg gctactacag cgtcaaaccc tccaagagc accactcaac cccgaccatc 2220
tcccttcca gctgcccagc cgaactgctg cctgcccgca agcagcgtgt gccacagggc 2280
atgtcttca ccaacatcta cagcaccctg agcggccagg gccgctctca cgaactcagg 2340
cagcgtttg tgcctggcct gggcagctcg tccatcgagc tttgtgagc ggaattccag 2400
agaggtctcc tcagcgacag cagctctctc ctggacacgc agtgtgacg cagcgtcagc 2460
agcagcgcca agcaggatgg ctatgtgca gttcgacaag ccagcaaggc ttctgcttcc 2520
tctctccacc actcccagtc ctctgtccc aactctgacc ccagtcgacc cctgacagcg 2580
cggatgcaga ctacgtcta aggatcacac accgcccgtg gggaccggcc agggagagag 2640
tcagggcacg tctgtgtgt ccaggacaga gggactctt gcagaggaca ccaagaattg 2700
cacttccag gacagcctcc cagcgcctct gccactgctt tctctgaaag ctctgatcaa 2760
gcaaaaatc gggctcccag gtgctgtgtg ccagaggtgg cgggtgggg agacagacag 2820
aggtgcccgc tgaagtgcct gtgcttagtg ctggacaccc ggtctcccgg ccttctctg 2880
gagcctctc taccacctgc tctgcccaca ggcacaagt gcagctataa cctctcttct 2940
atgaaactgc ggtccactct ctggtctctc tgtggctct accctctgct gaccagaagc 3000
tctactacc cctgtgctgt tctgcccata cagcctggg gagaagggga tgcagctctc 3060
ccagcactga cctgcccagc aaaccccggc tcccactgc tctctatagc ccatacctg 3120
gagcctgaca agccagaamt gcccttggct aaaggagcct ctctctcacc aggtggcgg 3180
ggagccccc ccaatttgt ttgggttttt ggttccatcc tcttgcagtt ctgtcttgg 3240
actgtatgcc cctgaactct ccggtgggac cgttccgctc agagcctggt gtaactggg 3300
gagggggga gaggggagcc ttgtctgagc gactcctcg ccgggtgtgc cctctctgg 3360

```

WO 02/072794

PCT/US02/09052

```

ctgtgtgacc ccagcctccc caccacactc ctgctttgtg taactctccc ctccccctca 3420
gcacaatcgg agttcatata agaagtgcgg gagctctctt ggtcagggtt ctctgaacac 3480
ttatggagag agtgctctct ggaagtggtg gcttttgaag gggctggagg gcaggtcttt 3540
aaagtggcga gactgccctt ctcaactgat aaacacaaga ccggcgatcc tgtcttcaat 3600
aaggctccac gagaagagag gaagtataac tacacctcaa cctctctagt caccacctga 3660
ataaatgttt agggacacta ctccaaaaaa aaaaa 3695

```

```

<210> 24
<211> 2403
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7503603CB1

```

```

<400> 24
caggaaatag tgacagtaag aatgacaagg gattaggact ggcttctct tataaataat 60
aaaaaccaaa gagaagtgc ttgagtctcc aggtttaaag gagagcaact agaagtcgtc 120
caaacacctg catctcataa ggagaagaaa agtccacctg gatcttcttt ctggactgag 180
atggatggag aggccacagt gaagcctgga gaacaaaagg aagtgtgag gagaggaaga 240
gaagtggact actccaggct cattgctggc actttacac aatctcaagt tcttcttleg 300
ccctccaca aaaaagacc catccagat ggctgtgga gggctctct cctccagga 360
cccctctccg ggcctggga aactcaggc ctctctgccc cctctgctgg cggagggaga 420
gctcctttgc agctccaaat tcattgaagg gctcaaggct ggtgtcagg gagcctggag 480
gagctgtcac catccagtc cattatgccc cctcatctgt caacaggcac cagaggaagt 540
actggtcccg tctggggccc ccaagatgga tetgcccagc cattgtgtcc accaaccagt 600
atactcacca tgcctatcgt gacctgtgg cctccacaga ctltccacag agaggtttgt 660
ttgtgttgag gctgtcccaa ctgtcccgg atgacatcg atgtaactc tgccgattg 720
gaagtgaana caactgctgt ttcttaagca tgaatctgac catctctgca ggtcccgcc 780
gcacctccc cacagccact ccagctgctg gggagctcac catgagatcc talggaacag 840
cgtctccagt ggcacaacaga tggccccag gaaccacca gaccttaga caggagacag 900
catgggacac agttgcttc actccaggaa ccagcaagac tacagctca gctgagggaa 960
gacgaacccc aggagcaacc aggcagcag ctccaggac aggcagctgg gcagaggtt 1020
ctgtcaaac accgtctccg attccagaga gtccacctc aaagacaga agcatgtcca 1080
atacaacaga aggtgtttgg gagggacca gaagctcgtt gacaacagg gctagagcca 1140
gcaaggacag gagggagatg acaactacca aggtctgatg gccaaaggag gacatagagg 1200
gggtcaggat agctcttgat gcagccaaaa aggtcctagc aaccattggg ccaccagctc 1260
tggctcaga aactttggcc tgggaatcc tcccacaagc aacgccagtt tctaagcaac 1320
aatctcagg ttccattgga gaaacaactc cagctgcagc catgtggacc ttgggaactc 1380
cagctcaga tgtgtggtc ttgggaactc cagctgcaga tgtgtggacc agcatggagg 1440
cagcatctgg ggaaggaagc gctgcagggg acctagatgc tgccactgga gacagaggtc 1500
cccagcaac actgagccag acccccggag taggacctgt gggacctctt ggcagggagt 1560
cctccgtgaa gcgtactttt ccagaagatg aaagcagctc tgggacctgt gctcctgtct 1620
ctaccatgct ggcctgtttt atgcttatgt ctctggttct atgcaagg agctctgga 1680
gaaggaggac ctctcaggag gcagaaaggg tcaccttaat tcagatgaca cattttctgg 1740
aagtgaacc ccaagcagac cagctgcccc atgtggaag aaagatgctc caggatgact 1800
ctctctgc tggggccagc ctgactgcc cagagagaga aatccaggac cctcagggac 1860
agagagatga actgctcagt tacatggga gaaggacaa gatcaaggc ctccaggacc 1920
ccagcctctt tccatcatcc ttctccacc tgtgggaaga gaagctgatg cagccggtgc 1980
tcccaccatg gaagaaagcc tggctctcct tggcccagc aaagctcaagc attatccagc 2040
tccaaagtg acaagatgac tcaaggaga ctccaagac agtgtatgaa aactggaag 2100
aggtcaacta ggaagacat gaaattcca ttctgaatg ttgaaaata gaaggagctt 2160
ccaatcagtg tggaaagtga caaatcccc atcaacactc ccagccttg ctggggctc 2220
ctttctgac tactgttagc actcagctc ccattcacat gtattatatt taagtgtacc 2280
agccttgcct tctcaagtag attctaaact ctttaagcc agtaattgca ttttatctgt 2340

```

WO 02/072794

PCT/US02/09052

ctcatgatgc cccagagaa ctccaactc agtaggaacc catttaatac ctgtgtctga 2400
ttg 2403

【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
19 September 2002 (19.09.2002)

PCT

(10) International Publication Number
WO 2002/072794 A3

- (51) International Patent Classification: C12N 9/00, 15/00, 15/63, 15/85, C07H 21/04, C07K 17/00
- (21) International Application Number: PCT/US2002/009052
- (22) International Filing Date: 12 March 2002 (12.03.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
- | | | |
|------------|--------------------------------|----|
| 60/275,249 | 12 March 2001 (12.03.2001) | US |
| 60/316,810 | 31 August 2001 (31.08.2001) | US |
| 60/523,977 | 21 September 2001 (21.09.2001) | US |
| 60/548,447 | 26 October 2001 (26.10.2001) | US |
| 60/543,880 | 2 November 2001 (02.11.2001) | 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): YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). XU, Yuming [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). THANGAVELU, Kavitha [IN/US]; 1950 Montecito Avenue #23, Mountain View, CA 94043 (US). WARREN, Bridget, A. [US/US]; 10130 Parkwood Drive # 2, Cupertino, CA 95014 (US). TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). DUGGAN, Brendan, M. [AU/US]; 243 Buena Vista Avenue # 306, Sunnyvale, CA 94086 (US). TRAN, Uyen, K. [US/US]; 2638 Mabury Square, San Jose, CA 95133 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (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, EC, 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, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, 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:
29 January 2004
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2002/072794 A3

(54) Title: IMMUNOGLOBULIN SUPERFAMILY PROTEINS

(57) Abstract: The invention provides human immunoglobulin superfamily proteins (IGSF) and polynucleotides which identify and encode IGSF. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of IGSF.

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT		International application No. PCT/US02/09052
A. CLASSIFICATION OF SUBJECT MATTER IPC(7) : C12N 9/00, 15/00, 15/63, 15/85; C07H 21/04; C07K 17/00 US CL : 435/69.1, 183, 320.1, 325; 536/23.1; 530/350 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) U.S. : 435/69.1, 183, 320.1, 325; 536/23.1; 530/350 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 practicable, search terms used) Please See Continuation Sheet		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A,P	DATABASE NCBI, NAGASE et al. KIAA 1867 protein. Accession No. BAB47496. June 2001.	1-79
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents:		
"A" document defining the general state of the art which is not considered to be of particular relevance	"**" later documents 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	
"B" earlier application or patent published on or after the international filing date	"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	
"L" document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"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	
"O" document referring to an oral disclosure, use, exhibition or other means	"Z" document member of the same patent family	
"P" document published prior to the international filing date but later than the priority date claimed		
Date of the actual completion of the international search 30 September 2003 (30.09.2003)	Date of mailing of the international search report 29 OCT 2003	
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703)305-3230	Authorized officer Sandra Wegert Telephone No. 703.308.0196	

INTERNATIONAL SEARCH REPORT

PCT/US02/09052

Continuation of B. FIELDS SEARCHED Item 3:
Medline, Biosis, USPAT, EPO, Derwent. Search Terms: Immunoglobulin, Ig, IGSFP, sapiens, human, KIAA*, NIEPH*, receptor,
RIKEN, and combinations. STIC search SEQ ID NO: 1-4 and interference. Inventors, PALM, EAST: Yue, H., Xu, Y.,
Thangavelu, K., Warren, B., Tang, Y., Duggan, B., Tran, U., Baughn, M., Honchell, C., Burford, N., Forsythe, I., Yang, J.,
Mason, P.

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 7/06	A 6 1 P 9/10	4 B 0 6 5
A 6 1 P 9/10	A 6 1 P 11/06	4 C 0 8 4
A 6 1 P 11/06	A 6 1 P 17/00	4 C 0 8 5
A 6 1 P 17/00	A 6 1 P 19/06	4 H 0 4 5
A 6 1 P 19/06	A 6 1 P 25/00	
A 6 1 P 25/00	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 25/28	
A 6 1 P 25/28	A 6 1 P 27/02	
A 6 1 P 27/02	A 6 1 P 29/00	1 0 1
A 6 1 P 29/00	A 6 1 P 31/04	
A 6 1 P 31/04	A 6 1 P 31/10	
A 6 1 P 31/10	A 6 1 P 31/12	
A 6 1 P 31/12	A 6 1 P 33/00	
A 6 1 P 33/00	A 6 1 P 35/00	
A 6 1 P 35/00	A 6 1 P 37/00	
A 6 1 P 37/00	A 6 1 P 43/00	1 0 5
A 6 1 P 43/00	C 0 7 K 14/47	
C 0 7 K 14/47	C 0 7 K 16/18	
C 0 7 K 16/18	C 1 2 M 1/00	A
C 1 2 M 1/00	C 1 2 N 1/15	
C 1 2 N 1/15	C 1 2 N 1/19	
C 1 2 N 1/19	C 1 2 N 1/21	
C 1 2 N 1/21	C 1 2 P 21/02	C
C 1 2 N 5/10	C 1 2 P 21/08	
C 1 2 P 21/02	C 1 2 Q 1/68	A
C 1 2 P 21/08	G 0 1 N 33/15	Z
C 1 2 Q 1/68	G 0 1 N 33/50	Z
G 0 1 N 33/15	G 0 1 N 33/53	D
G 0 1 N 33/50	G 0 1 N 33/53	M
G 0 1 N 33/53	G 0 1 N 37/00	1 0 2
G 0 1 N 37/00	C 1 2 N 5/00	A
	C 1 2 N 15/00	F
	A 6 1 K 37/02	

(31)優先権主張番号 60/348,447

(32)優先日 平成13年10月26日(2001.10.26)

(33)優先権主張国 米国(US)

(31)優先権主張番号 60/343,880

(32)優先日 平成13年11月2日(2001.11.2)

(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,ZM,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,EC,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,OM,PH,P L,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZM,ZW

- (72)発明者 スー、ユーミング
アメリカ合衆国カリフォルニア州9 4 0 4 0 ・マウンテンビュー・ウォルナットドライブ 1 7 3
9
- (72)発明者 サンガベル、カピサ
アメリカ合衆国カリフォルニア州9 4 0 4 3 ・マウンテンビュー・# 2 3 ・モンテシタアベニュー
1 9 5 0
- (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 5 0 1 4 ・クーペルティノー・# 2 ・パークウッドドライブ
1 0 1 3 0
- (72)発明者 タング、ワイ・トム
アメリカ合衆国カリフォルニア州9 5 1 1 8 ・サンノゼ・ランウィックコート 4 2 3 0
- (72)発明者 ダガン、ブレンダン・エム
アメリカ合衆国カリフォルニア州9 4 0 8 6 ・サニーバイル・# 3 0 6 ・ブエナビスタアベニュー
2 4 3
- (72)発明者 トラン、ユエン・ケイ
アメリカ合衆国カリフォルニア州9 5 1 3 3 ・サンノゼ・メイブリースクエア 2 6 3 8
- (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 ・サンレアンドロ・サンティアゴロード 1 4 2 4 4
- (72)発明者 ホンチェル、シンシア・ディー
アメリカ合衆国カリフォルニア州9 4 0 7 0 ・サンカルロス・# 2 0 3 ・ローレルストリート 4
0 0
- (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 フォーサイス、イアン・ジェイ
アメリカ合衆国カリフォルニア州9 4 0 6 1 ・レッドウッドシティ・ローブルアベニュー 3 0 8
- (72)発明者 ヤング、ジュンミング
アメリカ合衆国カリフォルニア州9 5 1 2 9 ・サンノゼ・パークレーン 7 1 2 5
- (72)発明者 メイソン、パトリシア・エム
アメリカ合衆国カリフォルニア州9 5 0 3 7 ・モーガンヒル・クラークレーン 3 6 0

F ターム(参考) 2G045 AA25 AA40 BB03 BB20 CB01 CB21 DA12 DA13 DA14 DA36
DA77 FB02 FB03
4B024 AA01 AA11 BA44 CA01 GA11 HA12 HA20
4B029 AA23 BB20 CC03
4B063 QA01 QA05 QA19 QQ02 QQ42 QQ53 QQ91 QQ96 QR08 QR32
QR42 QR48 QR55 QR62 QS33 QS34 QX01
4B064 AG01 AG20 AG26 AG27 CA10 CA19 CC24 CD20 CE13 DA01
DA13
4B065 AA93Y AB01 AC14 BA02 CA24 CA25 CA44 CA46
4C084 AA02 AA07 BA01 BA08 BA22 CA18 NA14 ZA02 ZA16 ZA33
ZA36 ZA45 ZA55 ZA59 ZA89 ZA94 ZA96 ZB05 ZB15 ZB21
ZB26 ZB31 ZB32 ZB33 ZB35 ZB37
4C085 HH01 KA03 KA04 LL01 LL07 LL09 LL13 LL15 LL18
4H045 AA11 AA20 AA30 BA10 CA40 DA50 DA75 DA76 EA20 EA50
FA74 GA26

专利名称(译)	免疫球蛋白超家族蛋白		
公开(公告)号	JP2004533222A	公开(公告)日	2004-11-04
申请号	JP2002571850	申请日	2002-03-12
[标]申请(专利权)人(译)	洞察Genomics公司		
申请(专利权)人(译)	洞察基因组公司		
[标]发明人	ユエヘンリー スーユーミング サンガベルカピサ ワレンブリジットエイ タングワイトム ダガンブレンダンエム トランユエンケイ ポーグンマライアール ホンチエルシンシアディー バーフォードニール フォーサイスイアンジェイ ヤングジュンミング メイソンパトリシアエム		
发明人	ユエ、ヘンリー スー、ユーミング サンガベル、カピサ ワレン、ブリジット・エイ タング、ワイ・トム ダガン、ブレンダン・エム トラン、ユエン・ケイ ポーグン、マライア・アール ホンチエル、シンシア・ディー バーフォード、ニール フォーサイス、イアン・ジェイ ヤング、ジュンミング メイソン、パトリシア・エム		
IPC分类号	G01N33/50 A61K38/00 A61K45/00 A61K49/00 A61P3/10 A61P7/06 A61P9/10 A61P11/06 A61P17/00 A61P19/06 A61P25/00 A61P25/16 A61P25/18 A61P25/28 A61P27/02 A61P29/00 A61P31/04 A61P31 /10 A61P31/12 A61P33/00 A61P35/00 A61P37/00 A61P43/00 C07K14/47 C07K14/705 C07K16/18 C12M1/00 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N15/09 C12P21/02 C12P21/08 C12Q1/68 G01N33/15 G01N33/53 G01N37/00		
CPC分类号	A01K2217/05 A61K38/00 A61K2039/505 A61P11/06 A61P17/00 A61P19/06 A61P25/00 A61P25/16 A61P25/18 A61P25/28 A61P27/02 A61P29/00 C07K14/70503		
FI分类号	C12N15/00.ZNA.A A61K45/00 A61K49/00 A61P3/10 A61P7/06 A61P9/10 A61P11/06 A61P17/00 A61P19/06 A61P25/00 A61P25/16 A61P25/18 A61P25/28 A61P27/02 A61P29/00.101 A61P31/04 A61P31/10 A61P31/12 A61P33/00 A61P35/00 A61P37/00 A61P43/00.105 C07K14/47 C07K16/18 C12M1/00.A 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 G01N37/00.102 C12N5/00.A C12N15/00.F 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/DA77 2G045/FB02 2G045/FB03 4B024/AA01 4B024/AA11 4B024/BA44 4B024/CA01 4B024/GA11 4B024/HA12 4B024/HA20 4B029/AA23 4B029/BB20 4B029		

/CC03 4B063/QA01 4B063/QA05 4B063/QA19 4B063/QQ02 4B063/QQ42 4B063/QQ53 4B063/QQ91
 4B063/QQ96 4B063/QR08 4B063/QR32 4B063/QR42 4B063/QR48 4B063/QR55 4B063/QR62 4B063
 /QS33 4B063/QS34 4B063/QX01 4B064/AG01 4B064/AG20 4B064/AG26 4B064/AG27 4B064/CA10
 4B064/CA19 4B064/CC24 4B064/CD20 4B064/CE13 4B064/DA01 4B064/DA13 4B065/AA93Y 4B065
 /AB01 4B065/AC14 4B065/BA02 4B065/CA24 4B065/CA25 4B065/CA44 4B065/CA46 4C084/AA02
 4C084/AA07 4C084/BA01 4C084/BA08 4C084/BA22 4C084/CA18 4C084/NA14 4C084/ZA02 4C084
 /ZA16 4C084/ZA33 4C084/ZA36 4C084/ZA45 4C084/ZA55 4C084/ZA59 4C084/ZA89 4C084/ZA94
 4C084/ZA96 4C084/ZB05 4C084/ZB15 4C084/ZB21 4C084/ZB26 4C084/ZB31 4C084/ZB32 4C084
 /ZB33 4C084/ZB35 4C084/ZB37 4C085/HH01 4C085/KA03 4C085/KA04 4C085/LL01 4C085/LL07
 4C085/LL09 4C085/LL13 4C085/LL15 4C085/LL18 4H045/AA11 4H045/AA20 4H045/AA30 4H045
 /BA10 4H045/CA40 4H045/DA50 4H045/DA75 4H045/DA76 4H045/EA20 4H045/EA50 4H045/FA74
 4H045/GA26

優先権
 60/275249 2001-03-12 US
 60/316810 2001-08-31 US
 60/323977 2001-09-21 US
 60/348447 2001-10-26 US
 60/343880 2001-11-02 US

外部リンク Espacenet

摘要(译)

本发明提供了鉴定和编码IGSFp的人免疫球蛋白超家族蛋白 (IGSFp) 和多核苷酸。本发明还提供表达载体, 宿主细胞, 抗体, 激动剂和拮抗剂。本发明还提供了用于诊断, 治疗或预防与IGSFp异常表达有关的疾病的方法。

		(P2004-5382ZZA)	
		(43) 公表日 平成16年11月4日(2004.11.4)	
(51) Int. Cl. ⁷	F I	ターマコード (参考)	
C 1 2 N 15/09	C 1 2 N 15/00	Z N A A	2 G 0 4 5
A 6 1 K 30/00	A 6 1 K 45/00		4 B 0 2 4
A 6 1 K 45/00	A 6 1 K 49/00		4 B 0 2 9
A 6 1 K 49/00	A 6 1 P 3/10		4 B 0 6 3
A 6 1 P 3/10	A 6 1 P 7/06		4 B 0 6 4
審査請求 未請求 予備審査請求 未請求 (全 245 頁) 最終頁に続く			
(21) 出願番号	特願2002-571850 (P2002-571850)	(71) 出願人	301005050
(60) (22) 出願日	平成14年3月12日 (2002.3.12)		インサイト・ゲノミクス・インコーポレ イテッド
(85) 翻訳文提出日	平成15年9月9日 (2003.9.9)		アメリカ合衆国カリフォルニア州9430 4・パロアルト・ボータードドライブ 31 60
(86) 国際出願番号	PCT/US2002/009052	(74) 代理人	100089266
(87) 国際公開番号	W02002/072794		弁理士 大島 剛一
(87) 国際公開日	平成14年9月19日 (2002.9.19)		ユエ、ヘンリー
(31) 優先権主張番号	60/275,249		アメリカ合衆国カリフォルニア州9408 7・サニーベイル・ルイスアベニュー 8 26
(32) 優先日	平成13年3月12日 (2001.3.12)	(72) 発明者	
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/316,810		
(32) 優先日	平成13年8月31日 (2001.8.31)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/323,977		
(32) 優先日	平成13年9月21日 (2001.9.21)		
(33) 優先権主張国	米国 (US)		

最終頁に続く