(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2004-535759
(P2004-535759A)
(43)公表日　平成16年12月2日(2004.12.2)

(51)Int.Cl.[7]	F I			テーマコード（参考）
C12N 15/09	C12N	15/00	A	2G045
A61K 45/00	A61K	45/00		4B024
A61K 49/00	A61K	49/00	A	4B029
A61P 1/04	A61P	1/04		4B063
A61P 1/16	A61P	1/16		4B064

審査請求 未請求　予備審査請求 有　（全 219 頁）　最終頁に続く

(21)出願番号　特願2002-543666(P2002-543666)
(86)(22)出願日　平成13年11月15日(2001.11.15)
(85)翻訳文提出日　平成15年5月12日(2003.5.12)
(86)国際出願番号　PCT/US2001/044974
(87)国際公開番号　WO2002/040671
(87)国際公開日　平成14年5月23日(2002.5.23)
(31)優先権主張番号　60/249,645
(32)優先日　平成12年11月16日(2000.11.16)
(33)優先権主張国　米国(US)

(71)出願人　301005050
インサイト・ゲノミックス・インコーポレイテッド
アメリカ合衆国カリフォルニア州９４３０４・パロアルト・ポータードライブ　３１６０
(74)代理人　100089266
弁理士　大島　陽一
(72)発明者　ボーグン、マライア・アール
アメリカ合衆国カリフォルニア州９４５７７・サンレアンドロ・サンティアゴロード　１４２４４
(72)発明者　リュ、デュング・アイナ・エム
アメリカ合衆国カリフォルニア州９５１２３・サンノゼ・コイドライブ　２３３
最終頁に続く

(54)【発明の名称】免疫グロブリンスーパファミリタンパク質

(57)【要約】
本発明はヒト免疫グロブリンスーパファミリタンパク質（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：１－６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からなる群から選択したアミノ酸配列と少なくとも９０％が同一であるような天然のアミノ酸配列を含むポリペプチド
（ｃ）ＳＥＱ ＩＤ ＮＯ：１－６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：１－６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
【請求項２】
ＳＥＱ ＩＤ ＮＯ：１－６からなる群から選択されたアミノ酸配列を含むことを特徴とする請求項１に記載の単離されたポリペプチド。
【請求項３】
請求項１のポリペプチドをコードする単離されたポリヌクレオチド。
【請求項４】
請求項２のポリペプチドをコードする単離されたポリヌクレオチド。
【請求項５】
ＳＥＱ ＩＤ ＮＯ：７－１２を有す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。
【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のポリペプチドを生産する方法であって、以下の過程を含む方法。
（ａ）前記ポリペプチドの発現に好適な条件下で、請求項１のポリペプチドをコードするポリヌクレオチドに機能的に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で形質転換された細胞を培養するステップと、
（ｂ） そのように発現した前記ポリペプチドを回収する過程とからなり、前記組換えポリヌクレオチドが、請求項１に記載の前記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ことを特徴とする方法。
【請求項１０】
前記ポリペプチドがＳＥＱ ＩＤ ＮＯ：１－６からなる群から選択したアミノ酸配列を有する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。
（ａ）ＳＥＱ ＩＤ ＮＯ：７－１２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：７－１２からなる群から選択したポリヌクレオチド配列と少なくとも９０％が同一であるような天然のポリヌクレオチド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
（ｅ）（ａ）～（ｄ）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有する少なくとも２０の連続したヌクレオチドを含むプローブを用いて前記サンプルをハイブリダイズする方法と、プローブと標的ポリヌクレオチドあるいはその断片の間でハイブリダイゼーション複合体が形成されるような条件下で、プローブは標的ポリヌクレオチドに特異的にハイブリダイズすることを特徴とする方法。
（ｂ）前記ハイブリダイゼーション複合体の存在・不存在を検出し、該複合体が存在する場合にはオプションでその量を検出する過程からなり、前記プローブと前記標的ポリヌクレオチドまたはその断片の間でハイブリダイゼーション複合体が形成されるような条件下で、前記プローブが前記標的ポリヌクレオチドに特異的にハイブリダイズすることを特徴とする方法。
【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） 前記標的ポリヌクレオチドまたはその断片の存在・不存在を検出し、該標的ポリヌクレオチドまたはその断片が存在する場合にはオプションでその量を検出する過程を含むことを特徴とする方法。
【請求項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：１－６からなる群から選択されたアミノ酸配列を含むことを特徴とする請求項１７の組成物。
【請求項１９】
機能性ＩＧＳＦＰの発現低下に関連する疾患又は病状を治療する方法であって、そのような治療を必要とする患者に対して請求項１７に記載の組成物を投与する過程を含むことを特徴とする方法。
【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有効性を確認するために化合物をスクリーニングする方法であって、
（ａ） 請求項１のポリペプチドを含むサンプルを化合物に曝すステップと、
（ｂ）前記サンプルにおいてアゴニスト活性を検出するステップとを含むことを特徴とするスクリーニング方法。
【請求項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組成物。
【請求項２２】
機能性ＩＧＳＦＰの発現低下に関連する疾患又は病状を治療する方法であって、そのような治療を必要とする患者に対して請求項２１に記載の組成物を投与する過程を含むことを特徴とする方法。
【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有効性を確認するために化合物をスクリーニングする方法であって、
（ａ） 請求項１のポリペプチドを含むサンプルを化合物に曝すステップと、
（ｂ）前記サンプルにおいてアゴニスト活性を検出するステップとを含むことを特徴とするスクリーニング方法。
【請求項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】
機能性ＩＧＳＦＰの過剰発現に関連する疾患又は病状の治療方法であって、そのような治療を必要とする患者に対して請求項２４に記載の組成物を投与する過程を含むことを特徴とする方法。
【請求項２６】
請求項１に記載のポリペプチドに特異結合する化合物をスクリーニングする方法であって、
（ａ）適切な条件下で請求項１に記載のポリペプチドを少なくとも１つの試験化合物に混合させる過程と、
（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ステップを含むことを特徴とする方法。
【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に記載のポリペプチドの活性が許容された条件下で、請求項１に記載のポリペプチドを少なくとも１つの試験化合物に混合する過程と、
（ｂ）請求項１に記載のポリペプチドの活性を試験化合物の存在下で算定する過程と、
（ｃ）試験化合物の存在下での請求項１に記載のポリペプチドの活性を、試験化合物の不存在下での請求項１に記載のポリペプチドの活性と比較する過程を含み、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の配列を含む標的ポリヌクレオチドの発現を変化させるのに効果的な化合物をスクリーニングする方法であって、
（ａ）前記標的ポリヌクレオチドの発現に好適な条件下で、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、
（ｂ）前記標的ポリヌクレオチドの変異発現を検出する過程と、
（ｃ）可変量の前記化合物の存在下と前記化合物の不存在下で、前記標的ポリヌクレオチドの発現を比較する過程とを含むことを特徴とする方法。
【請求項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含む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ゼーションが、前記プローブと前記生物学的サンプルの標的ポリヌクレオチドとの間で特異的なハイブリダイゼーション複合体が形成される条件下で行われ、前記標的ポリヌクレオチドが、請求項１２に記載のポリヌクレオチドのポリヌクレオチド配列またはその断片を含むポリヌクレオチドである、前記過程と、
（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前記ハイブリタイゼーション複合体の量を、処理されていない生物学的サンプル中の前記ハイブリタイゼーション複合体の量と比較す

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

【請求項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）前記抗体が前記ポリペプチドに結合し、抗体とポリペプチドとの複合体が形成されるのに適した条件下で、前記生物学的サンプルを請求項１１に記載の抗体と結合する過程と

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

【請求項３１】

前記抗体が、

（ａ）キメラ抗体

（ｂ）単鎖抗体

（ｃ）Ｆａｂ断片

（ｄ）Ｆ（ａｂ’）$_2$ 断片

（ｅ）ヒト化抗体のいずれかであることを特徴とする請求項１１に記載の抗体。

【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】

被検者のＩＧＳＦＰ の発現に関連する病状又は疾患の診断方法であって、請求項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】

被検者のＩＧＳＦＰ の発現に関連する病状又は疾患の診断方法であって、請求項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：１－６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：１－６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に特異結合するようなポリクローナル抗体を同定する過程とを含むことを特徴とする方法。

【請求項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好適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製造する方法であって、

（ａ）抗体反応を誘発する条件下で、ＳＥＱ ＩＤ ＮＯ：１－６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）前記抗体産出細胞を不死化の細胞と融合して、モノクローナル抗体を産出するハイブリドーマ細胞を形成する過程と、

（ｄ）前記ハイブリドーマ細胞を培養する過程と、

（ｅ）ＳＥＱ ＩＤ ＮＯ：１－６からなる群から選択したアミノ酸配列を有するポリペ

プチドに特異結合するような前記培養モノクローナル抗体から単離する過程とを含むことを特徴とする方法。
【請求項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好適なキャリアを含む組成物。
【請求項４２】
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：１－６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をサンプル中から検出する方法であって、
（ａ）前記抗体と前記ポリペプチドの特異結合を許容する条件下で、サンプルを用いて請求項１１に記載の抗体をインキュベートする過程と、
（ｂ）特異結合を検出する過程とを含み、ＳＥＱ ＩＤ ＮＯ：１－６を有する群から選択したアミノ酸配列を有するポリペプチドがサンプル中に存在することを該特異結合が示唆することを特徴とする方法。
【請求項４５】
ＳＥＱ ＩＤ ＮＯ：１－６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を精製する方法であって、
（ａ）前記抗体と前記ポリペプチドの特異結合を許容する条件下で、サンプルを用いて請求項１１に記載の抗体をインキュベートする過程と、
（ｂ） 前記サンプルから前記抗体を分離し、ＳＥＱ ＩＤ ＮＯ：１－６からなる群から選択したアミノ酸配列を有する精製ポリペプチドを得る過程とを含むことを特徴とする方法。
【請求項４６】
マイクロアレイの少なくとも一つの要素が請求項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成する方法であって、
（ａ）サンプルのポリヌクレオチドを標識する過程と、
（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、および
（ｃ）サンプル中のポリヌクレオチドの発現量を定量する過程を含む。
【請求項４８】
固体基質上の物理的に異なった位置に付着した異なったヌクレオチド分子からなるアレイであって、該ヌクレオチド分子の少なくとも一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ズすることが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、また、該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ようなアレイ。
【請求項４９】
請求項４８に記載のアレイであって、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の少なくとも３０の近接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って、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の少なくとも６０の近接するヌクレオチドに完全に相補的であるアレイ。
【請求項５１】
請求項４８に記載のアレイであって、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】
マイクロアレイである請求項４８に記載のアレイ。
【請求項５３】
前記の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さらに含む請求項４８に記載のアレイ。
【請求項５４】
請求項４８のアレイであって、リンカーが前記ヌクレオチド分子の少なくとも一つを前記固体基板に結合するアレイ。
【請求項５５】
請求項４８に記載のアレイであって、基板上の物理的に異なった各々の位置が複数のヌクレオチド分子を含み、またその複数のヌクレオチド分子が任意の一つの物理的に異なった位置で同じ配列を有し、基板上の物理的に異なった各々の位置が基板上の別の物理的に異なった位置のヌクレオチド分子の配列とは異なった配列を有するヌクレオチド分子を含むようなアレイ。
【請求項５６】
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５７】
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５８】
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５９】
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６０】
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。
【請求項６１】
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６２】
ＳＥＱ ＩＤ ＮＯ：７のポリヌクレオチド配列を有する請求項１２に記載のポリヌクレオチド。
【請求項６３】
ＳＥＱ ＩＤ ＮＯ：８のポリヌクレオチド配列を有する請求項１２に記載のポリヌクレオチド。
【請求項６４】
ＳＥＱ ＩＤ ＮＯ：９のポリヌクレオチド配列を有する請求項１２に記載のポリヌクレオチド。
【請求項６５】
ＳＥＱ ＩＤ ＮＯ：１０のポリヌクレオチド配列を有する請求項１２に記載のポリヌクレオチド。
【請求項６６】
ＳＥＱ ＩＤ ＮＯ：１１のポリヌクレオチド配列を有する請求項１２に記載のポリヌクレオチド。
【請求項６７】
ＳＥＱ ＩＤ ＮＯ：１２のポリヌクレオチド配列を有する請求項１２に記載のポリヌクレオチド。

【発明の詳細な説明】
【０００１】
（技術分野）
本発明は、免疫グロブリンスーパファミリ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疾患、神経系疾患、発生または発達障害、筋肉疾患および癌を含む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ファミリ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。
【０００２】
（発明の背景）
認識、接着または結合のような機能を媒介する大部分の細胞表面や可溶性分子は共通の進化的前駆体から進化している（すなわち、これらのタンパク質は構造的相同性を有する）。同様の機能を有する免疫系外の多数の分子もまた、この同じ進化的前駆体から由来している。これらの分子は免疫グロブリン（Ｉｇ）スーパファミリのメンバーとして分類される。Ｉｇスーパファミリのメンバーであるためのタンパク質の判定基準は、Ｉｇ可変性様（Ｖ）ドメインか、またはＩｇ定常性様（Ｃ）ドメインのどちらかに相同的な、長さが７０～１１０ のアミノ酸残基の領域であるＩｇドメインを一つ以上有することである。Ｉｇスーパファミリのメンバーには、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）が含まれる。
【０００３】
Ｉｇドメイン（Ｖ および Ｃ） は、ポリペプチドに免疫グロブリン（または抗体）フォールド（折畳み）と呼ばれる球状三次構造を与える保存されたアミノ酸残基の領域である。免疫グロブリン（または抗体）フォールドはほぼ平行になった２層のβシートからなる。保存されたシステイン残基は、長さ５５～７５ のアミノ酸残基の鎖内ジスルフィド結合ループを形成し、２層のβシートをつなぐ。各々のβシートは、５～１０アミノ酸残基の逆平行β鎖を３～４つ有する。β鎖内のアミノ酸残基の疎水性および親水性相互作用によって、Ｉｇフォールド（鎖の内向き部分のアミノ酸残基に疎水性であり、また、鎖の外向き部分のアミノ酸残基に親水性である）は安定化される。ＶドメインはＣドメインより長いポリペプチドからなり、Ｉｇフォールドにおいてもう一対のβ鎖を有する。
【０００４】
Ｉｇスーパファミリ遺伝子の一貫した特徴は、Ｉｇドメインの各々の配列が一つのエキソンによってコードされることである。このスーパファミリは、細胞間相互作用の媒介に関与している単一の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。次に、このスーパファミリの新しいメンバーがエキソンと遺伝子の複製によって生じた。現代のＩｇスーパファミリタンパク質は異なった数のＶドメイン および／または Ｃ ドメインを含んでいる。このスーパファミリの他の進化的特徴は、ＤＮＡ再配列を生じる能力で、このファミリの抗原受容体メンバーによって保持されているユニークな特徴である。
【０００５】
Ｉｇスーパファミリの多くのメンバーは細胞外Ｉｇ ドメインを持つ膜内在性タンパク質である。膜貫通ドメインおよび細胞質テイルの疎水性アミノ酸残基は極めて多様であり、Ｉｇ ファミリメンバー間で殆ど、あるいは全く相同性がないか、または、既知のシグナル伝達構造に殆ど、あるいは全く相同性がない。この一般的なスーパファミリの記述に関して、例外がある。例えば、ＰＤＧＦＲ の細胞質テイルはチロシンキナーゼ活性を有する。さらに、Ｔｈｙ－１ は胸腺細胞およびＴ細胞に存在する糖タンパク質である。このタンパク質には細胞質テイルはないが、そのかわり、グリコホスファチジルイノシトール共有結合によって原形質膜に結合している。
【０００６】
多くのＩｇ スーパファミリタンパク質の他の共通の特徴は、これらの分子の機能に不可

欠であるＩｇ ドメイン間の相互作用である。 多量体タンパク質のＩｇ ドメイン間の相互作用は同種親和性か、または異種親和性（すなわち、同じＩｇドメイン間、または異なったＩｇドメイン間）のどちらかであり得る。抗体はＩｇドメイン間で同種親和性および異種親和性相互作用の両方を持つ多量体タンパク質である。重鎖の定常領域の対形成によって抗体のＦｃ領域が形成され、軽鎖と重鎖の可変領域の対形成によって抗体の抗原結合部位が形成される。異種親和性相互作用はまた、異なった分子のＩｇドメイン間でも生じる。これらの相互作用によって、免疫系および発達中の、または成熟した神経系において重要な細胞間相互作用のための細胞間接着が生じる。（Ａｂｂａｓ，Ａ．Ｋ．他（１９９１）Ｃｅｌｌｕｌａｒ ａｎｄ Ｍｏｌｅｃｕｌａｒ Ｉｍｍｕｎｏｌｏｇｙ，Ｗ．Ｂ．Ｓａｕｎｄｅｒｓ Ｃｏｍｐａｎｙ，Ｐｈｉｌａｄｅｌｐｈｉａ ＰＡ，１４２－１４５ページに概説されている。）

【０００７】

抗体

抗体はＩｇスーパファミリの多量体メンバーで、これはＢ細胞の表面に発現されるか、またはＢ細胞によって循環系に分泌される。抗体は血液中や、他の細胞外液中で外来性抗原に結合して、これを中和する。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二つの同じポリペプチド重鎖（Ｈ鎖）と二つの同じポリペプチド軽鎖（Ｌ鎖）からなる四量体である。この配列によって、抗体分子が特徴的なＹ型を形成する。抗体はＨ鎖組成に基づいて分類される。五つの抗体のクラスである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、どちらかが対として、いずれのＨ鎖対に会合し得る。血液循環中に存在する抗体の最も一般的なクラスの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。

【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か、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、特定のクラスのＨ鎖間またはＬ鎖間でほとんど同じである。可変領域は、Ｈ鎖とＬ鎖の両方で約１１０のアミノ酸からなる。

しかし、可変領域のアミノ酸配列は、特定のクラスのＨ鎖間またはＬ鎖間で異なる。Ｈ鎖またはＬ鎖の可変領域内に広範囲配列多様性の三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一緒になって抗原認識部位を形成する（Ａｌｂｅｒｔｓ，他（１９９４）Ｍｏｌｅｃｕｌａｒ Ｂｉｏｌｏｇｙ ｏｆ ｔｈｅ Ｃｅｌｌ，Ｇａｒｌａｎｄ Ｐｕｂｌｉｓｈｉｎｇ，Ｎｅｗ Ｙｏｒｋ ＮＹ，１２０６－１２１３ ページと １２１６－１２１７ページ参照）。

【０００９】

Ｈ鎖およびＬ鎖は両方共、Ｉｇスーパファミリ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鎖は四つのＩｇドメインを含んでおり、そのうちの三つは定常領域内に発生し、一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、そのうちの一つは定常領域内に発生し、他の一つは可変領域内に発生する。

【００１０】

免疫系は体内に入る外来性分子を認識し、それに応答することができる。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最大限の範囲の抗体で武装している必要がある。このような抗体の多様性は可変領域と定常領域をコードする遺伝子区分の体細胞性再配列によって作られる。これらの遺伝子区分は、各々の遺伝子区分が隣接する極めて保存されたＤＮＡ配列間で生じる部位特異的組換えによって連結される。何百という異なった遺伝子区分があるため、何百万という固有の遺伝子が組み合わされて作成され得る。その上、これらの区分の不正確な結合とこれらの区分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さらに寄与している。

【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時の神経ネットワークの確立に役割を果たしている（Ｕｙｅｍｕｒａ 他（１９９６）Ｅｓｓａｙｓ Ｂｉｏｃｈｅｍ．３１：３７－４８；Ｂｒｕｍｍｅｎｄｏｒｆ 及びＲａｔｈｊｅｎ（１９９６）Ｃｕｒｒ．Ｏｐｉｎ．Ｎｅｕｒｏｂｉｏｌ．６：５８４－５９３）。ＮＣＡＰは神経細胞の遊走、細胞接着、神経突起の生成、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、遺伝性神経障害の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）。
【００１２】
ＮＣＡＰは少なくとも一つの免疫グロブリンの定常または可変ドメインを持っている（Ｕｙｅｍｕｒａ 他、前出）。これらは一般に膜貫通型ドメインやグリコシルホスファチジルイノシトール（ＧＰＩ）アンカーによって原形質膜に結合している。ＧＰＩ 結合はＧＰＩホスホリパーゼＣによって切断され得る。大部分のＮＣＡＰは一つ以上の免疫グロブリンドメイン、膜貫通ドメインおよび細胞内領域から構成されている。多くのＮＣＡＰは、共有結合で付いたオリゴ糖、グルクロン酸および硫酸などの翻訳後修飾を含む。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）およびアヘン剤結合細胞接着分子（ＯＢＣＡＭ）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チロシンキナーゼおよび上皮成長因子様、ｓｅｍａおよびＰＳＩ（プレキシン、セマホリンおよび インテグリン） ドメインなどの他のモチーフ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）やニューロトロピン（ＮＴ４）等の神経成長因子のＴｒｋ受容体、グリア成長因子ＩＩ（ＧＧＦＩＩ）やアセチルコリン受容体誘導因子（ＡＲＩＡ）等のＮｅｕ分化因子、およびセマフォリンＢやコラプシン等のセマフォリン／コラプシンファミリ、並びにプレキシンのようなセマホリン／コラプシンファミリメンバーの受容体（プレキシンについては以下参照）を含んでいる。
【００１３】
あるＮＣＡＰサブファミリ（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アンカーによっ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する（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の発現は発生的に調節されている。胚および出産後の脳と比べて、成人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ＩＩＩ型ドメインを含む。ＳＤＫはショウジョウバエの眼の発達において細胞間相互作用に関与している（Ｎｇｕｙｅｎ，Ｄ．Ｎ．Ｔ．他 （１９９７） Ｄｅｖｅｌｏｐｍｅｎｔ １２４：３３０３）。

【００１５】

シナプス膜の糖タンパク質

特殊化された細胞接合部が細胞間の接触部分で生じる。これらの細胞接合部の内に、細胞間の化学的シグナルや電気的シグナルの通過を媒介する伝達接合部がある。中枢神経系において、ニューロンとニューロンの間の伝達接合部はシナプス接合部として知られている。これらはシナプス前ニューロンおよびシナプス後ニューロンの膜および細胞骨格から構成されている。シナプス膜（ＳＭ）分画およびシナプス後密度（ＰＳＤ）分画のような生化学的に単離されたシナプス亜分画に存在するいくつかの糖タンパク質が同定され、その機能が確立されている。一つの例は、Na^+/K^+－ＡＴＰアーゼのβ２ サブユニットとして同定されたＳＭ糖タンパク質のｇｐ５０である。

【００１６】

２つの糖タンパク質、ｇｐ６５ および ｇｐ５５はラット前脳から作製されたシナプス膜の主要構成成分である。これらはＩｇスーパファミリのメンバーであり、各々３個および２個のＩｇドメインを含む。Ｉｇ スーパファミリのメンバーとして、これらのタンパク質の可能な機能として、シナプス接合部における接着相互作用の媒介が提唱されている（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、その結果、細胞の凝集をもたらす細胞外糖タンパク質の遍在的なファミリから構成されている（Ｄｒｉｃｋａｍｅｒ，Ｋ．ａｎｄ Ｔａｙｌｏｒ，Ｍ．Ｅ．（１９９３） Ａｎｎｕ．Ｒｅｖ．Ｃｅｌｌ Ｂｉｏｌ．９：２３７－２６４の概説を参照）。この機能は免疫応答の活性化にとって特に重要である。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】

シアル酸結合性Ｉｇ様レクチン（ＳＩＧＬＥＣ） は、糖タンパク質および糖脂質のシアル酸に結合するＩｇスーパファミリのメンバーである。ＳＩＧＬＥＣにはシアロアドヘジン、ＣＤ２２、ＣＤ３３、ミエリン結合糖タンパク質 （ＭＡＧ）、 ＳＩＧＬＥＣ－５、ＳＩＧＬＥＣ－６、ＳＩＧＬＥＣ－７および ＳＩＧＬＥＣ－８がある。ＳＩＧＬＥＣ の細胞外領域には膜から遠位のＶ－セットドメインがあり、これに異なった数のＣ２セットドメインが続く。シアル酸結合ドメインはＶセットドメインにマップされる。神経系のみに発現されるＭＡＧ を除いて、大部分のＳＩＧＬＥＣは造血系細胞の固有のサブセットに発現される。例えば、ＳＩＧＬＥＣ－８ は、多核白血球（顆粒球）の一形態である好酸球のみに発現される （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）はタンパク質間相互作用に関与する。哺乳類神経細胞ロイシンリッチリピートタンパク質（ＮＬＬＲ－１ および ＮＬＬＲ－２）、ショウジョウバエのコネクチン、ｓｌｉｔ、ｃｈａｏｐｉｎおよびｔｏｌｌのようなＬＲＲＰはすべて神経細胞の発達において役割を果たす。ＬＲＲＰの細胞外領域には異なった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およびフィブロネクチン型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セット および Ｃ２ セットに加えて、ＩＰＴ／ＴＩＧ（プレキシンと転写因子が共有する免疫グロブリン様フォールド）と呼ばれるＤセットの免疫グロブリン様ドメインがある。ＩＰＴ／ＴＩＧ を含むタンパク質にはプレキシン、ＭＥＴ／ＲＯＮ／ＳＥＡ（肝細胞成長因子受容体ファミリ）および転写因子ＸＣｏｅ２（アフリカツメガエルの初代神経細胞の特定化に関与するＣｏｌ／Ｏｌｆ１／ＥＢＦ ファミリの転写因子）が含まれる（Ｂｏｒｋ，Ｐ．他（１９９９）Ｔｒｅｎｄｓ Ｂｉｏｃｈｅｍ．Ｓｃｉ．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およびＶＥＳＰＲ のようなプレキシンは軸索誘導を制御する神経細胞セマホリン受容体であることがわかっている（Ｗｉｎｂｅｒｇ Ｍ．Ｌ．他（１９９８）Ｃｅｌｌ ９５：９０３９１６）。

【００２１】
新規の免疫グロブリンスーパファミリタンパク質、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、免疫系疾患、神経系疾患、発生または発達障害、筋疾患および癌を含む細胞増殖異常の診断・治療・予防において有用であり、また、免疫グロブリンスーパファミリ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も有用である。

【００２２】
（発明の要約）
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」、および「ＩＧＳＦＰ－５」と呼ぶ、免疫グロブリンスーパファミリタンパク質である精製されたポリペプチドを提供する。或る実施態様において本発明は、（ａ）ＳＥＱ ＩＤ ＮＯ：１－６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：１－６を有する群から選択したアミノ酸配列と少なくとも９０％が同一性の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：１－６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、および（ｄ）ＳＥＱ ＩＤ ＮＯ：１－６を有する群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片を含む群から選択した単離されたポリペプチドを提供する。一実施態様では、ＳＥＱ ＩＤ ＮＯ：１－６のアミノ酸配列を含む単離されたポリペプチドを提供する。

【００２３】
また、本発明は（ａ）ＳＥＱ ＩＤ ＮＯ：１－６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：１－６を有する群から選択したアミノ酸配列と少なくとも９０％が同一性のある天然のアミノ酸配列からなるポリペプチド、（ｃ）ＳＥＱ ＩＤ ＮＯ：１－６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、および（ｄ）ＳＥＱ ＩＤ ＮＯ：１－６を有する群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片、を含む群から選択されたポリペプチドをコードする単離されたポリヌクレオチドを提供する。一実施態様では、ポリヌクレオチドはＳＥＱ ＩＤ ＮＯ：１－６を有する群から選択したポリペプチドをコードする。

別の実施態様では、ポリヌクレオチドはＳＥＱ ＩＤ ＮＯ：７－１２を有する群から選択される。

【００２４】

本発明は更に、（ａ）ＳＥＱ ＩＤ ＮＯ：１－６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：１－６を有する群から選択したアミノ酸配列と少なくとも９０％の同一性のある天然のアミノ酸配列からなるポリペプチド、（ｃ）ＳＥＱ ＩＤ ＮＯ：１－６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：１－６を有する群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片を有する群から選択したポリペプチドをコードするようなポリヌクレオチドと機能的に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。別の実施態様では、本発明は組換えポリヌクレオチドを含む遺伝形質転換体を提供する。

【００２５】

更に本発明は、（ａ）ＳＥＱ ＩＤ ＮＯ：１－６からなる群から選択されたアミノ酸配列からなるポリペプチドと、（ｂ）ＳＥＱ ＩＤ ＮＯ：１－６からなる群から選択されたアミノ酸配列と９０％以上の配列同一性を有する天然のアミノ酸配列からなるポリペプチドと、（ｃ）ＳＥＱ ＩＤ ＮＯ：１－６からなる群から選択されたアミノ酸配列を有するポリペプチドの生物学的に活性な断片と、（ｄ）ＳＥＱ ＩＤ ＮＯ：１－６とからなる群から選択されたアミノ酸配列を有するポリペプチドの免疫原性断片とで構成される群から選択されたポリペプチドを製造する方法を提供する。製造方法は、（ａ）組換えポリヌクレオチドを用いて形質転換した細胞をポリペプチドの発現に適した条件下で培養する過程と、（ｂ）そのように発現したポリペプチドを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。

【００２６】

本発明は更に、（ａ）ＳＥＱ ＩＤ ＮＯ：１－６からなる群から選択したアミノ酸配列を有するポリペプチド、（ｂ）ＳＥＱ ＩＤ ＮＯ：１－６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：１－６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：１－６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。

【００２７】

本発明は更に、（ａ）ＳＥＱ ＩＤ ＮＯ：７－１２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、（ｂ）ＳＥＱ ＩＤ ＮＯ：７－１２からなる群から選択したポリヌクレオチド配列と少なくとも９０％の同一性を有する天然のポリヌクレオチド配列を含むポリヌクレオチド、（ｃ）（ａ）に相補的なポリヌクレオチド配列、（ｄ）（ｂ）に相補的なポリヌクレオチド配列、および（ｅ）（ａ）～（ｄ）のＲＮＡ等価物からなる群から選択された単離されたポリヌクレオチドを提供する。一実施態様では、ポリヌクレオチドは少なくとも６０の連続したヌクレオチドを有する。

【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：７－１２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：７－１２からなる群から選択したポリヌクレオチド配列と少なくとも９０％の同一性を有する天然のポリヌクレオチド配列を含むポリヌクレオチド、（ｃ）（ａ）に相補的なポリヌクレオチド、（ｄ）（ｂ）に相補的なポリヌクレオチド配列、および（ｅ）（ａ）～（ｄ）のＲＮＡ等価物からなる群から選択されたポリヌクレオチド配列を含む。検出方法は、（ａ）サンプル中の標的ポリヌクレオチドに相補的な配列を含む少なくとも２０の連続したヌクレオチド

を含むプローブを用いて該サンプルをハイブリダイズする過程と、（ｂ）ハイブリダイゼーション複合体の存在・不存在を検出し、複合体が存在する場合にはオプションでその量を検出する過程からなり、プローブと標的ポリヌクレオチドあるいはその断片の間でハイブリダイゼーション複合体が形成されるような条件下で、プローブは標的ポリヌクレオチドに特異的にハイブリダイズする。一実施態様では、プローブは少なくとも６０の連続したヌクレオチドを含む。

【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：７－１２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：７－１２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）増幅され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。

【００３０】

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し、そのポリペプチドは、（ａ）ＳＥＱ ＩＤ ＮＯ：１－６を有する群から選択したアミノ酸配列を有するポリペプチド、（ｂ）ＳＥＱ ＩＤ ＮＯ：１－６を有する群から選択したアミノ酸配列と少なくとも９０％の同一性の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：１－６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：１－６を有する群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片を含む群から選択される。一実施例では、ＳＥＱ ＩＤ ＮＯ：１－６からなる群から選択されたアミノ酸配列を含む組成物を提供する。 更に、本発明は、患者にこの組成物を投与することを含む、機能的ＩＧＳＦＰ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。

【００３１】

本発明はまた、（ａ）ＳＥＱ ＩＤ ＮＯ：１－６からなる群から選択したアミノ酸配列を有するポリペプチド、（ｂ）ＳＥＱ ＩＤ ＮＯ：１－６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有するポリペプチド、（ｃ）ＳＥＱ ＩＤ ＮＯ：１－６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：１－６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。また別の実施態様において、本発明は、機能的ＩＧＳＦＰの発現低下に関連する疾患または病状を治療する方法を提供し、この方法はそのような治療を必要とする患者に対して組成物を投与する過程を含む。

【００３２】

本発明は更に、（ａ）ＳＥＱ ＩＤ ＮＯ：１－６を有する群から選択したアミノ酸配列を有するポリペプチド、（ｂ）ＳＥＱ ＩＤ ＮＯ：１－６を有する群から選択したアミノ酸配列と少なくとも９０％の同一性の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：１－６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：１－６を有する群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片を含む群から選択されたポリペプチドのアンタゴニストとしての有効性を確認するために化合物をスクリーニングする方法を提供

する。 スクリーニング方法は、（ａ）ポリペプチドを含むサンプルを化合物に曝す過程と、（ｂ）サンプル中のアンタゴニスト活性を検出する過程とを含む。一実施態様で本発明は、この方法によって同定したアンタゴニスト化合物と薬剤として好適な賦形剤とを含む組成物を提供する。更なる別法では、本発明は、このような治療を必要とする患者へのこの組成物の投与を含む、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する。

【００３３】

本発明は更に、（ａ）ＳＥＱ ＩＤ ＮＯ：１－６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：１－６を有する群から選択したアミノ酸配列の生物学的活性断片、または（ｄ）ＳＥＱ ＩＤ ＮＯ：１－６からなる群から選択したアミノ酸配列の免疫原性断片を含む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つの試験化合物に混合する過程と、（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。

【００３４】

本発明は更に、（ａ）ＳＥＱ ＩＤ ＮＯ：１－６を有する群から選択したアミノ酸配列を有するポリペプチド、（ｂ）ＳＥＱ ＩＤ ＮＯ：１－６を有する群から選択したアミノ酸配列と少なくとも９０％の同一性を有する天然のアミノ酸配列を有するポリペプチド、（ｃ）ＳＥＱ ＩＤ ＮＯ：１－６を有する群から選択したアミノ酸配列の生物学的活性断片、または（ｄ）ＳＥＱ ＩＤ ＮＯ：１－６を有する群から選択したアミノ酸配列の免疫原性断片を含む群から選択し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る条件下で、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。

【００３５】

更に本発明は、ＳＥＱ ＩＤ ＮＯ：７－１２からなる群から選択されたポリヌクレオチドを含む標的ポリヌクレオチドの発現を変化させるのに有効な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）この標的ポリヌクレオチドを含むサンプルを化合物に曝露するステップと、（ｂ）この標的ポリヌクレオチドの発現の変化を検出するステップとを含む。

【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：７－１２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：７－１２からなる群から選択したポリヌクレオチド配列と少なくとも９０％の同一性を有する天然のポリヌクレオチド配列を含むポリヌクレオチド、（ｉｉｉ）（ｉ）のポリヌクレオチドに相補的な配列を有するポリヌクレオチド、（ｉｖ）（ｉｉ）のポリヌクレオチドに相補的なポリヌクレオチド、（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：７－１２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：７－１２からなる群から選択したポリヌクレオチド配列と少なくとも９０％の同一性を有する天然のポリヌクレオ

チド配列を含むポリヌクレオチド、（ｉｉｉ）（ｉ）のポリヌクレオチドに相補的な配列を有するポリヌクレオチド、（ｉｖ）（ｉｉ）のポリヌクレオチドに相補的なポリヌクレオチド、および（ｖ）（ｉ）～（ｉｖ）のＲＮＡ等価物からなる群から選択する。或いは、標的ポリヌクレオチドは、上記（ｉ）～（ｖ）からなる群から選択したポリヌクレオチド配列の断片と、（ｃ）ハイブリダイゼーション複合体の量を定量する過程と、（ｄ）処理した生物学的サンプルのハイブリダイゼーション複合体の量を、非処理の生物学的サンプルのハイブリダイゼーション複合体の量と比較する過程を含み、処理した生物学的サンプルのハイブリダイゼーション複合体の量の差は、試験化合物の毒性を示す。

【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」と記されている場合にはそのような宿主細胞が複数あることもあり、「或る抗体」と記されている場合には単数または複数の抗体、及び、当業者に公知の抗体の等価物等についても言及しているのである。

【００３９】

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

【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】

用語「アゴニスト」は、ＩＧＳＦＰの生物学的活性を強めたり、模倣する分子を指す。このアゴニストは、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 の活性を調節するタンパク質、核酸、糖質、小分子、任意の他の化合物や組成物を含み得る。

【００４２】

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

【００４３】

ＩＧＳＦＰをコードする「変異（ａｌｔｅｒｅｄ）」核酸配列は、種々のヌクレオチドを

欠失、挿入または置換する核酸配列を有し、ＩＧＳＦＰと同一またはＩＧＳＦＰの機能的特徴を少なくとも１つ有するポリペプチドを産出する。この定義には、ＩＧＳＦＰをコードするポリヌクレオチド配列の正常な染色体の遺伝子座以外の位置での対立遺伝子変異配列との不適当或いは予期しないハイブリダイゼーション、並びに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、及び／または両親媒性についての類似性に基づいて成され得る。例えば、負に帯電したアミノ酸にはアスパラギン酸及びグルタミン酸があり、正に帯電したアミノ酸にはリジン及びアルギニンがある。親水性値が近似している非荷電極性側鎖を有するアミノ酸には、アスパラギンとグルタミン、セリンとトレオニンがある。親水性値が近似している非荷電側鎖を有するアミノ酸には、ロイシンとイソロイシンとバリン、グリシンとアラニン、フェニルアラニンとチロシンがある。

【００４４】

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

【００４５】

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

【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。アンタゴニストとしては、抗体などのタンパク質、核酸、糖質、小分子またはその他の任意の化合物や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エレメントに作用することによって、ＩＧＳＦＰ の活性を調節する。

【００４７】

「抗体」の語は、抗原決定基と結合することができる、そのままの免疫グロブリン分子やその断片、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、そのままのポリペプチド、または関心のある小ペプチドを含む断片を用いて作製可能である。動物（マウス、ラット、ウサギ等）を免疫化するために用いるポリペプチドまたはオリゴペプチドは、ＲＮＡの翻訳、または化学合成によって得られるポリペプチドまたはオリゴペプチドに由来し得るもので、好みに応じて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。その結合ペプチドは、動物を免疫化するために用いる。

【００４８】

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

【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。アプタマーの組成は２本鎖か、または一本鎖であり、デオキシリボヌクレオチド、リボヌクレオチド、ヌクレオチド誘導体、或い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構成要素は修飾された糖基（例えば、リボヌクレオチドの２’－ＯＨ 基が２’－Ｆ または ２’－NH_2 で置換されている）を有することが可能で、これらの糖基は、例えば、ヌクレアーゼに対する耐性あるいは血中でのより長い生存期間などの望ましい性質を改善する。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等の他の分子に抱合させることができる。アプタマーは、例えば、架橋剤の光活性化によって各々のリガンドと特異的に架橋させることができる（例えば、Ｂｒｏｄｙ，Ｅ．Ｎ．及びＬ．Ｇｏｌｄ（２０００）Ｊ．Ｂｉｏｔｅｃｈｎｏｌ．７４：５－１３参照）。

【００５０】

用語「イントラマー（ｉｎｔｒａｍｅｒ）」はｉｎ ｖｉｖｏで発現されるアプタマーを指す。例えば、ワクシニアウイルス性ＲＮＡ 発現系は白血球の細胞質において高レベルで特定のＲＮＡアプタマーを発現させるのに使われている（Ｂｌｉｎｄ，Ｍ．他（１９９９）Ｐｒｏｃ．Ｎａｔｌ Ａｃａｄ．Ｓｃｉ．ＵＳＡ ９６：３６０６－３６１０）。

【００５１】

用語「スピーゲルマー（ｓｐｉｅｇｅｌｍｅｒ）」は Ｌ－ＤＮＡ、Ｌ－ＲＮＡ、または、他の左旋性ヌクレオチド誘導体あるいはヌクレオチド様分子を含むアプタマーを指す。左旋性ヌクレオチドを含むアプタマーは、右旋性のヌクレオチドを含む基質に通常作用する天然の酵素による分解に耐性がある。

【００５２】

「アンチセンス」の語は、特定の核酸配列の「センス」（コーディング）鎖と塩基対を形成することが可能な任意の成分を指す。アンチセンス組成物には、ＤＮＡ、ＲＮＡ、ペプチド核酸（ＰＮＡ）、ホスホロチオネート、メチルホスホネートまたはベンジルホスホネートのような修飾されたバックボーンリンケージを持つオリゴヌクレオチド、２’－メトキシエチル糖あるいは２’－メトキシエトキシ糖のような修飾された糖基を持つオリゴヌクレオチド、あるいは、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のような修飾された塩基を持つオリゴヌクレオチドが含まれ得る。アンチセンス分子は、化学合成または転写を含む任意の方法で製造することができる。相補的アンチセンス分子は、ひとたび細胞に導入されたら、細胞が形成した天然の核酸配列と塩基対を形成し、転写または翻訳を阻止する二重鎖を形成する。「負」若しくは「マイナス（－）」の語がある参照ＤＮＡ分子のアンチセンス鎖を、「正」若しくは「プラス（＋）」がある参照ＤＮＡ分子のセンス鎖を指すことがある。

【００５３】

「生物学的に活性」の語は、天然分子の構造的機能、調節機能または生化学的機能を有するタンパク質を指す。同様に、用語「免疫学的に活性」または「免疫原性」は、天然或いは組換え体のＩＧＳＦＰ、合成のＩＧＳＦＰまたはそれらの任意のオリゴペプチドが、適切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）」または「相補性」の語は、塩基対形成によってアニーリングする２つの一本鎖核酸の間の関係を指す。例えば、配列「５’Ａ－Ｇ－Ｔ３’」は、相補配列「３’Ｔ－Ｃ－Ａ５’」に結合する。

【００５５】

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

【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、ＸＬ－ＰＣＲキット（Ａｐｐｌｉｅｄ Ｂｉｏｓｙｓｔｅｍｓ，Ｆｏｓｔｅｒ Ｃｉｔｙ ＣＡ）を用いて５’及び／または３’の方向に伸長され、再度シークエ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）か、あるいはＰｈｒａｐ（Ｕｎｉｖｅｒｓｉｔｙ ｏｆ Ｗａｓｈｉｎｇｔｏｎ，Ｓｅａｔｔｌｅ ＷＡ）等の断片構築用のコンピュータプログラムを用いて１つ或いはそれ以上の重複するｃＤＮＡやＥＳＴ、またはゲノムＤＮＡ断片から構築された核酸配列を指す。伸長及び構築の両方を行ってコンセンサス配列を作製する配列もある。

【００５７】

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

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

【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βシートやαヘリックス構造、（ｂ）置換部位における分子の電荷または疎水性、及び／または（ｃ）側鎖の大部分が保持される。
【００５９】
「欠失」は、結果的に１個若しくは数個のアミノ酸残基またはヌクレオチドが失われてなくなるようなアミノ酸またはヌクレオチド配列における変化を指す。
【００６０】
「誘導体」の語は、ポリペプチドまたはポリヌクレオチドの化学修飾を指す。例えば、アルキル基、アシル基、ヒドロキシル基またはアミノ基による水素の置換は、ポリヌクレオチド配列の化学修飾に含まれ得る。ポリヌクレオチド誘導体は、天然分子の生物学的または免疫学的機能を少なくとも１つは保持しているポリペプチドをコードする。ポリペプチド誘導体は、グリコシル化、ポリエチレングリコール化（ｐｅｇｙｌａｔｉｏｎ）、或いは任意の同様なプロセスであって誘導起源のポリペプチドから少なくとも１つの生物学的若しくは免疫学的機能を保持するプロセスによって、修飾されたポリペプチドである。

【００６１】
「検出可能な標識」は、測定可能な信号を生成することができ、ポリヌクレオチドまたはポリペプチドに共有結合または非共有結合するようなレポーター分子または酵素を指す。
【００６２】
「示差発現」は少なくとも２つの異なったサンプルを比較することによって決められる、増加（上方調節）、あるいは減少（下方調節）、または遺伝子発現またはタンパク発現の欠損を指す。このような比較は例えば、治療後サンプルと未治療のサンプルまたは病態のサンプルと正常サンプルの間で行われ得る。
【００６３】
「エキソンシャフリング」は、異なるコード領域の組換えを指す（エキソン）。エキソンはコードされたタンパク質の構造的或いは機能的ドメインを表しているので、新たなタンパク質は安定的な下部構造の新規な再組合せにより構築される。それによって新たなタンパク質機能の進化の加速が可能となる。
【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）と配列は同一であるが、長さがより短いものを指す。断片は、画定された配列の全長から１ヌクレオチド／アミノ酸残基を差し引いた長さよりも短い長さを有し得る。例えば或る断片は、５～１０００の連続したヌクレオチドまたはアミノ酸残基を有し得る。プローブ、プライマー、抗原、治療用分子として、或いはその他の目的のために用いられる断片は、少なくとも５、１０、１５、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長さであり得る。断片は、分子の特定領域から選択的に選択し得る。例えば、ポリペプチド断片は、所定の配列に示すような最初の２５０または５００アミノ酸（またはポリペプチドの最初の２５％または５０％）から選択された或る長さの連続したアミノ酸を有し得る。これらの長さは明らかに例として挙げているものであり、本発明の実施例では、配列表、表及び図面を含む明細書に裏付けされた任意の長さであってよい。
【００６５】
ＳＥＱ　ＩＤ　ＮＯ：７－１２ の断片は、例えば、この断片を得たゲノム内の他の配列とは異なる、ＳＥＱ　ＩＤ　ＮＯ：７－１２ を明確に同定する固有のポリヌクレオチド配列の領域を含む。ＳＥＱ　ＮＯ　ＩＤ：７－１２の断片は、例えば、ハイブリダイゼーション及び増幅技術において、或いは関連するポリヌクレオチド配列からＳＥＱ　ＮＯ　ＩＤ：７－１２を区別する類似の方法において有用である。ＳＥＱ　ＩＤ　ＮＯ：７－１２の断片の正確な長さ及び断片に対応するＳＥＱ　ＩＤ　ＮＯ：７－１２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。
【００６６】
ＳＥＱ　ＩＤ　ＮＯ：１－６のある断片はＳＥＱ　ＩＤ　ＮＯ：７－１２のある断片によってコードされている。ＳＥＱ　ＩＤ　ＮＯ：１－６の断片はＳＥＱ　ＩＤ　ＮＯ：１－６を特異的に同定する固有のアミノ酸配列の領域を含む。例えば、ＳＥＱ　ＩＤ　ＮＯ：１－６の断片は、ＳＥＱ　ＩＤ　ＮＯ：１－６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：１－６のある断片の正確な長さとその断片に対応するＳＥＱ　ＩＤ　ＮＯ：１－６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。
【００６８】
「相同性」の語は、配列類似性即ち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間で互換可能な配列同一性である。

【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少なくとも２つ以上のポリヌクレオチド配列間で一致する残基の割合を意味する。このようなアルゴリズムは、２配列間のアラインメントを最適化するために比較する配列において、標準化された再現性のある方法でギャップを挿入するので、２つの配列をより有意に比較できる。
【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 ソフトウエアパッケージ（一組の分子生物学的分析プログラム）（ＤＮＡＳＴＡＲ，Ｍａｄｉｓｏｎ ＷＩ）の一部である。このＣＬＵＳＴＡＬ Ｖは、Ｈｉｇｇｉｎｓ，Ｄ．Ｇ．及びＰ．Ｍ．Ｓｈａｒｐ（１９８９）ＣＡＢＩＯＳ ５：１５１－１５３、Ｈｉｇｇｉｎｓ，Ｄ．Ｇ．他（１９９２）ＣＡＢＩＯＳ ８：１８９－１９１に記載されている。ポリヌクレオチド配列をペアワイズでアラインメントする際のデフォルトパラメータは、Ｋｔｕｐｌｅ＝２、ｇａｐ ｐｅｎａｌｔｙ＝５、ｗｉｎｄｏｗ＝４、「ｄｉａｇｏｎａｌｓ ｓａｖｅｄ」＝４と設定する。「重み付けされた」残基重み付け表が、デフォルトとして選択された。ＣＬＵＳＴＡＬ Ｖは、アラインメントされたポリヌクレオチド配列対間の「類似率」として一致率を報告する。
【００７１】
或いは、米国国立バイオテクノロジー情報センター（ＮＣＢＩ）のＢａｓｉｃ Ｌｏｃａｌ Ａｌｉｇｎｍｅｎｔ Ｓｅａｒｃｈ Ｔｏｏｌ（ＢＬＡＳＴ）が一般的に用いられ、且つ、無料で入手可能な配列比較アルゴリズム一式を提供している（Ａｌｔｓｃｈｕｌ，Ｓ．Ｆ．他（１９９０）Ｊ．Ｍｏｌ．Ｂｉｏｌ．２１５：４０３－４１０）。このアルゴリズムは、幾つかの情報源から入手可能であり、メリーランド州ベセスダにあるＮＣＢＩ及びインターネット（ｈｔｔｐ：／／ｗｗｗ．ｎｃｂｉ．ｎｌｍ．ｎｉｈ．ｇｏｖ／ＢＬＡＳＴ／）からも入手可能である。ＢＬＡＳＴソフトウェア一式には様々な配列分析プログラムが含まれており、既知のポリヌクレオチド配列を種々のデータベースから得た別のポリヌクレオチド配列とアラインメントする「ｂｌａｓｔｎ」もその１つである。その他にも、２つのヌクレオチド配列をペアワイズで直接比較するために用いる「ＢＬＡＳＴ ２ Ｓｅｑｕｅｎｃｅｓ」と称されるツールも利用可能である。「ＢＬＡＳＴ ２ Ｓｅｑｕｅｎｃｅｓ」は、ｈｔｔｐ：／／ｗｗｗ．ｎｃｂｉ．ｎｌｍ．ｎｉｈ．ｇｏｖ／ｇｏｒｆ／ｂｌ２．ｈｔｍｌにアクセスして対話形式で利用することが可能である。「ＢＬＡＳＴ ２ Ｓｅｑｕｅｎｃｅｓ」ツールは、ｂｌａｓｔｎ と ｂｌａｓｔｐ（後述）の両方に用いることができる。ＢＬＡＳＴプログラムは、一般的には、ギャップ及びデフォルト設定に設定された他のパラメータと共に用いる。例えば、２つのヌクレオチド配列を比較するために、デフォルトパラメータとして設定された「ＢＬＡＳＴ ２ Ｓｅｑｕｅｎｃｅｓ」ツールＶｅｒｓｉｏｎ ２．０．１２（２０００年４月２１日）を用いてｂｌａｓｔｎを実行してもよい。デフォルトパラメータの設定例を以下に示す。

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

【００７２】
一致率は、完全に画定された（例えば特定の配列番号で画定された）配列長さと比較して測定し得る。 或いは、より短い長さ、例えばより大きな画定された配列から得られた断

片（例えば少なくとも２０、３０、４０、５０、７０、１００または２００の連続したヌクレオチドの断片）の長さと比較して一致率を測定してもよい。ここに挙げた長さは単なる例示的なものに過ぎず、表、図及び配列リストを含めた本明細書に記載された配列に裏付けられた任意の配列長さの断片を用いて、一致率を測定し得る長さを説明し得ることを理解されたい。

【００７３】

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。

【００７４】

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

【００７５】

ポリペプチド配列間の一致率は、ＭＥＧＡＬＩＧＮ ｖｅｒｓｉｏｎ ３．１２ｅ配列アラインメントプログラムに組込まれているようなＣＬＵＳＴＡＬ Ｖアルゴリズムのデフォ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でアラインメントする際のデフォルトパラメータは、Ｋｔｕｐｌｅ＝１、ｇａｐ ｐｅｎａｌｔｙ＝３、ｗｉｎｄｏｗ＝５、「ｄｉａｇｏｎａｌｓ ｓａｖｅｄ」＝５と設定する。デフォルトの残基重み付け表としてＰＡＭ２５０マトリクスを選択する。ポリヌクレオチドアラインメントと同様に、ＣＬＵＳＴＡＬ Ｖは、アラインメントされたポリペプチド配列対間の「類似率」として一致率を報告する。

【００７６】

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をする場合、ある者は、デフォルトパラメータで設定された「ＢＬＡＳＴ ２ Ｓｅｑｕｅｎｃｅｓ」ツールＶｅｒｓｉｏｎ ２．０．１２（Ａｐｒ－２１－２０００）でｂｌａｓｔｐを使用するであろう。デフォルトパラメータの設定例を以下に示す。

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

【００７７】

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

【００７８】

「ヒト人工染色体（ＨＡＣ）」は、約６ｋｂ（キロベース）～１０Ｍｂのサイズの ＤＮＡ配列を含み得る、染色体の複製、分離及び維持に必要な全てのエレメントを含む直鎖状の

微小染色体である。

【００７９】

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

【００８０】

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

【００８１】

一般に、ハイブリダイゼーションのストリンジェンシーは或る程度、洗浄ステップを実行する温度を基準にして表すことができる。このような洗浄温度は通常、所定のイオン強度及びｐＨにおける特定の配列の融点（Ｔｍ）より約５～２０℃低くなるように選択する。このＴｍは、所定のイオン強度及びｐＨの下で、完全に一致するプローブに標的配列の５０％がハイブリダイズする温度である。Ｔｍ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 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。

【００８２】

本発明のポリヌクレオチド間の高いストリンジェンシーのハイブリダイゼーションでは、約０．２ｘ ＳＳＣ及び約０．１％のＳＤＳの存在の下、約６８℃で１時間の洗浄過程を含む。別法では、６５℃、６０℃、５５℃、または４２℃の温度で行う。ＳＳＣ濃度は、約０．１％のＳＤＳ存在下で、約０．１～２×ＳＳＣの範囲で変化し得る。通常は、ブロッキング剤を用いて非特異ハイブリダイゼーションを阻止する。このようなブロッキング剤には、例えば、約１００～２００μｇ／ｍｌの変性サケ精子ＤＮＡがある。例えばＲＮＡとＤＮＡのハイブリダイゼーションのような特定条件下では、有機溶剤、例えば約３５～５０％ｖ／ｖの濃度のホルムアミドを用いることもできる。洗浄条件の有用なバリエーションは、当業者には自明であろう。 ハイブリダイゼーションは、特に高ストリンジェント条件下では、ヌクレオチド間の進化的な類似性を示唆し得る。このような類似性は、ヌクレオチド及びヌクレオチドにコードされるポリペプチドに対する類似の役割を強く示唆している。

【００８３】

「ハイブリダイゼーション複合体」の語は、相補的塩基対間の水素結合の形成力によって２つの核酸配列間に形成された複合体を指す。ハイブリダイゼーション複合体は、溶解状態で形成し得る（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。

【００８４】

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

【００８５】

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

【００８６】

用語「免疫原性断片」は、例えば哺乳動物などの生きている動物に導入すると、免疫反応を引き起こすＩＧＳＦＰ のポリペプチド断片またはオリゴペプチド断片を指す。用語「免疫原性断片」はまた、本明細書で開示するまたは当分野で周知のあらゆる抗体生産方法に有用なＩＧＳＦＰ のポリペプチド断片またはオリゴペプチド断片を含む。

【００８７】

「マイクロアレイ」の語は、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他の化合物の構成を指す。

【００８８】

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

【００８９】

「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における亢進または低下が起こる。

【００９０】

「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、また、一本鎖または二本鎖であり、ペプチド核酸（ＰＮＡ）や、任意のＤＮＡ様またはＲＮＡ様物質に対してセンス鎖またはアンチセンス鎖を表し得るようなゲノム起源または合成起源のＤＮＡまたはＲＮＡを指す。

【００９１】

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

【００９２】

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

【００９３】

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り得る。

【００９４】

「プローブ」とは、同一配列或いは対立遺伝子核酸配列、関連する核酸配列の検出に用いる、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」は、短い核酸、通常はＤＮＡオリゴヌクレオチドであり、相補的塩基対を形成することで標的ポリヌクレオチドにアニーリングされ得る。プライマーは次に、ＤＮＡポリメラーゼ酵素によって標的ＤＮＡ鎖に沿って伸長し得る。プライマー対は、例えばポリメラーゼ連鎖反応（ＰＣＲ）による核酸配列の増幅（及び同定）に用い得る。

【００９５】

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

【００９６】

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．ら，（１９８７）Ｃｕｒｒｅｎｔ Ｐｒｏｔｏｃｏｌｓ ｉｎ Ｍｏｌｅｃｕｌａｒ Ｂｉｏｌｏｇｙ，Ｇｒｅｅｎｅ Ｐｕｂｉ．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。

【００９７】

プライマーとして用いるオリゴヌクレオチドの選択は、そのような目的のために本技術分野でよく知られているソフトウェアを用いて行う。例えばＯＬＩＧＯ ４．０６ソフトウェアは、各１００ヌクレオチドまでのＰＣＲプライマー対の選択に有用であり、オリゴヌクレオチド及び最大５，０００ヌクレオチドまでの大きめのポリヌクレオチドであって３２キロベースまでのインプットポリヌクレオチド配列から得たものを分析するのにも有用である。類似のプライマー選択プログラムには、拡張能力のための追加機能が組込まれている。例えば、ＰｒｉｍＯＵプライマー選択プログラム（テキサス州ダラスにあるテキサス大学南西部医療センターのゲノムセンターから一般向けに入手可能）は、メガベース配列から特定のプライマーを選択することが可能であり、従ってゲノム全体の範囲でプライマーを設計するのに有用である。Ｐｒｉｍｅｒ３プライマー選択プログラム（マサチューセッツ州ケンブリッジのＷｈｉｔｅｈｅａｄ Ｉｎｓｔｉｔｕｔｅ／ＭＩＴゲノム研究センターから一般向けに入手可能）ではユーザーが「ミスプライミング・ライブラリ」をインプットすることができ、ここでプライマー結合部位として避けたい配列はユーザーが指定する。Ｐｒｉｍｅｒ３は特に、マイクロアレイのためのオリゴヌクレオチドの選択に有用である。（後二者のプライマー選択プログラムのソースコードは、各自のソースから得てユーザー固有のニーズを満たすように変更してもよい。）ＰｒｉｍｅｒＧｅｎプログラム（英国ケンブリッジ市の英国ヒトゲノムマッピングプロジェクト－リソースセンターから一般向けに入手可能）は、多数の配列アラインメントに基づいてプライマーを設計し、それによって、アラインメントされた核酸配列の最大保存領域または最小保存領域の何れかとハイブリダイズするようなプライマーの選択を可能にする。従って、このプログラムは、固有な、および保存されたオリゴヌクレオチド及びポリヌクレオチドの断片の同定に

有用である。上記選択方法のいずれかによって同定したオリゴヌクレオチド及びポリヌクレオチドの断片は、ハイブリダイゼーション技術において、例えばＰＣＲまたはシークエ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。

【００９８】

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

【００９９】

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。

【０１００】

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

【０１０１】

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

【０１０２】

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

【０１０３】

「サンプル」の語は、その最も広い意味で用いられる。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、細胞から単離された染色体や細胞内小器官、または膜である細胞からの抽出物と、細胞と、溶液中に存在するか、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。

【０１０４】

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

【０１０５】

「実質上精製された」の語は、自然環境から取り除かれ、或いは単離または分離された核酸またはアミノ酸配列であって、少なくとも約６０％、好ましくは少なくとも約７５％、

最も好ましいのは少なくとも約９０％の〈自然に会合している〉他の構成成分を含まないものを指す。

【０１０６】

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

【０１０７】

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

【０１０８】

「転写イメージ」または「発現プロフィール」は、所与の時間、条件での特定の細胞タイプまたは組織による遺伝子発現の集合的パターンを指す。

【０１０９】

「形質転換」は、外来性のＤＮＡが受入細胞に入り込むプロセスを表す。形質転換は、本技術分野で知られている種々の方法に従って自然条件または人工条件下で生じ得るものであり、外来性の核酸配列を原核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またはウイルス感染、電気穿孔法（エレクトロポレーション）、熱ショック、リポフェクション及び微粒子銃を用いる方法がある。「形質転換された細胞」には、導入されたＤＮＡが自己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、限られた時間に導入ＤＮＡ若しくは導入ＲＮＡを発現する、一過的に形質転換された細胞も含まれる。

【０１１０】

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

【０１１１】

特定の核酸配列の「変異体」は、核酸配列１本全部の長さに対して特定の核酸配列と少なくとも４０％の相同性を有する核酸配列であると定義する。 その際、デフォルトパラメータに設定した「ＢＬＡＳＴ ２ Ｓｅｑｕｅｎｃｅｓ」ツールＶｅｒｓｉｏｎ ２．０．９（１９９９年５月７日）を用いてｂｌａｓｔｎを実行する。このような核酸対は、所定の長さに対して、例えば少なくとも５０％、６０％、７０％、８０％、８５％、９０％、９１％、９２％、９３％、９４％、９５％、９６％、９７％、９８％、９９％またはそれ以上の相同性を示し得る。或る変異体は、例えば「対立遺伝子」変異体（前述）、「スプライス」変異体、「種」変異体または「多型性」変異体として説明し得る。スプライス変異体は参照分子とかなりの相同性を有し得るが、ｍＲＮＡプロセッシング中のエキソンの選択的スプライシングによって通常、ヌクレオチドがより多くまたはより少数の塩基を有することになる。対応するポリペプチドは、追加機能ドメインを有するか或いは参照分子

に存在するドメインが欠落していることがある。種変異体は、種相互に異なるポリヌクレオチド配列である。結果的に生じるポリペプチドは通常、相互にかなりのアミノ酸相同性を有する。多型性変異体は、与えられた種の個体間で特定の遺伝子のポリヌクレオチド配列が異なる。また、多型性変異体は、１つのヌクレオチド塩基によってポリヌクレオチド配列が変化する「１塩基多型」（ＳＮＰ）を含み得る。ＳＮＰの存在は、例えば特定の個体群、病状または病状性向を示し得る。

【０１１２】

特定のポリペプチド配列の「変異体」は、ポリペプチド配列の１本の長さ全体で特定のポリペプチド配列に対して少なくとも４０％の相同性を有するポリペプチド配列として画定される。 ここで、デフォルトパラメータに設定した「ＢＬＡＳＴ ２ 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。

【０１１３】

（発明）

本発明は、新規のヒト免疫グロブリンスーパファミリタンパク質（ＩＧＳＦＰ）および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疾患、神経系疾患、発生または発達障害、筋肉疾患および癌を含む細胞増殖異常の診断、治療、及び予防に関する。

【０１１４】

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

【０１１５】

表２は、ＧｅｎＢａｎｋタンパク質（ｇｅｎｐｅｐｔ）データベースに対するＢＬＡＳＴ分析によって同定されたような、本発明のポリペプチドとの相同性を有する配列を示している。列１および列２はそれぞれ、本発明の各ポリペプチドに対するポリペプチド配列識別番号（ポリペプチド 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、ＧｅｎＢａｎｋの最も近い相同体のＧｅｎＢａｎｋの識別番号（ＧｅｎＢａｎｋ ＩＤ ＮＯ ：）を示す。列４は、各ポリペプチドとその相同体との間の一致を表す確率スコアを示す。列５は、ＧｅｎＢａｎｋ相同体の注釈を示し、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。

【０１１６】

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、およびモチーフを含むアミノ酸残基を示す。列７は、タンパク質の構造／機能の分析のための分析方法を示し、該当箇所にはさらに分析方法に利用した検索可能なデータベースを示す。

【０１１７】
表２及び表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、それらの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ファミリタンパク質であることを確立している。例えば、ＳＥＱ ＩＤ ＮＯ：１はマウスのｐｕｎｃ（推定神経細胞接着タンパク質）（ＧｅｎＢａｎｋ ＩＤ ｇ３０６８５９２）と８７％の同一性を有することがＢａｓｉｃ Ｌｏｃａｌ Ａｌｉｇｎｍｅｎｔ Ｓｅａｒｃｈ Ｔｏｏｌ（ＢＬＡＳＴ）によって示された。（表２参照）ＢＬＡＳＴ確率スコアは０．０であるが、これは観測されたポリペプチド配列が偶然に得られる確率を示す。ＳＥＱ ＩＤ ＮＯ：１はまた、免疫グロブリンドメインとフィブロネクチンタイプＩＩＩ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 解析からのデータはＳＥＱ ＩＤ ＮＯ：１が神経細胞接着タンパク質であるというさらに確証的な証拠を提供する。配列識別番号２－６は、同様にして解析され注釈された。ＳＥＱ ＩＤ ＮＯ：１－６の解析のためのアルゴリズム及びパラメータが表７で記述されている。

【０１１８】
表４に示すように、本発明の完全長ポリヌクレオチド配列は、ｃＤＮＡ配列、またはゲノムＤＮＡ由来のコード（エキソン）配列、或いはこれらの２種類の配列のあらゆる組み合わせを用いて組み立てた。列１および列２はそれぞれ、本発明の各ポリヌクレオチドのポリヌクレオチド配列識別番号（ポリヌクレオチド ＳＥＱ ＩＤ ＮＯ ：）およびそれに対応するＩｎｃｙｔｅ ポリヌクレオチドコンセンサス配列番号（Ｉｎｃｙｔｅ ポリヌクレオチド ＩＤ）を示す。列３は、各ポリヌクレオチド配列の長さ（塩基対単位）を示す。列４は、例えば、ＳＥＱ ＩＤ ＮＯ：７－１２を同定する、或いはＳＥＱ ＩＤ ＮＯ：７－１２と、関連するポリヌクレオチド配列とを区別するハイブリダイゼーションまたは増幅技術に有用なポリヌクレオチド配列の断片を示す。列５はｃＤＮＡ配列、ゲノムＤＮＡから予想されたコード配列（エキソン）及び／またはｃＤＮＡ及びゲノムＤＮＡを共に有する配列集合に対応する識別番号を示している。これらの配列は、本発明の完全長ポリヌクレオチド配列を構築するのに用いた。表４の列６および列７はそれぞれ、列５の配列に対応するｃＤＮＡ配列および／またはゲノム配列の開始ヌクレオチド（５’）位置および終了ヌクレオチド（３’）位置を示す。

【０１１９】
表４の列５の識別番号は、特に例えばＩｎｃｙｔｅ ｃＤＮＡとそれに対応するｃＤＮＡライブラリを指す場合もある。例えば、７１８４８２１Ｈ１はＩｎｃｙｔｅｃＤＮＡ配列の識別番号であり、ＢＯＮＲＦＥＣ０１はそれが由来するｃＤＮＡライブラリの識別番号である。ｃＤＮＡライブラリが示されていないＩｎｃｙｔｅ ｃＤＮＡは、プールされているｃＤＮＡライブラリ（例えば、７１５５３４７０Ｖ１）に由来する。または、列５の識別番号は、完全長のポリヌクレオチド配列の構築に用いたＧｅｎＢａｎｋのｃＤＮＡすなわちＥＳＴ（例えば、ｇ２９９３９９６）の識別番号の場合もある。さらに、列５の識別番号はＥＮＳＥＭＢＬ（Ｔｈｅ Ｓａｎｇｅｒ Ｃｅｎｔｒｅ，Ｃａｍｂｒｉｄｇｅ，ＵＫ）データベース（すなわち、「ＥＮＳＴ」と命名された配列）から由来した配列を示す。あるいは、列５の識別番号はＮＣＢＩ ＲｅｆＳｅｑ ヌクレオチド配列記録データベース（すなわち、「ＮＭ」または「ＮＴ」と命名された配列）またはＮＣＢＩＲｅｆＳｅｑ タンパク質配列記録（すなわち、「ＮＰ」と命名された配列）から由来している。または列５の識別番号は、「エキソンスティッチング（ｅｘｏｎ－ｓｔｉｔｃｈｉｎｇ）」アルゴリズムにより結び合わせたｃＤＮＡ及びＧｅｎｓｃａｎ予想エキソンの両方の集合を指す場合もある。例えば、ＦＬ＿ＸＸＸＸＸＸ＿Ｎ$_1$＿Ｎ$_2$＿ＹＹＹＹＹ＿Ｎ$_3$＿Ｎ$_4$ は「スティッチ」配列を表しており、ＸＸＸＸＸＸはアルゴリズムが適用された配列のクラスターの識別番号であり、また、ＹＹＹＹＹ はアルゴリズムによって出された予測の番号である。Ｎ$_{1,2,3..}$ が存在する場合は、解析している時に手動で編集された可能性のある特定のエキソンを示す（実施例５を参照）。または列５の識別番号は、

「エキソンストレッチング（ｅｘｏｎ－ｓｔｒｅｔｃｈｉｎｇ）」アルゴリズムにより結び合わせたエキソンの集合を指す。例えば、ＦＬＸＸＸＸＸＸ＿ｇＡＡＡＡＡＡ＿ｇＢＢＢＢＢＢ＿１＿Ｎ は「ストレッチ」配列の識別番号である。ここでＸＸＸＸＸＸはＩｎｃｙｔｅプロジェクト識別番号、ｇＡＡＡＡＡＡは「エキソンストレッチング」アルゴリズムを適用したヒトゲノム配列のＧｅｎＢａｎｋ識別番号、ｇＢＢＢＢＢＢ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 識別番号である。（実施例５を参照。）ＲｅｆＳｅｑ 配列が「エキソンストレッチング」アルゴリズムのタンパク質相同体として使われた場合は、ＲｅｆＳｅｑ 識別子（「ＮＭ 」、「ＮＰ」または「 ＮＴ」と表示された）はＧｅｎＢａｎｋ 識別子（すなわち、ｇＢＢＢＢＢ）の代わりに使われる場合もある。

【０１２０】

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

接頭コード	解析のタイプまたはプログラムの例
GNN、GFG、ENST	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からのエキソン予測
GBI	手動で編集されたゲノム配列の解析
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	ゲノムへのEST配列のマッピングからの完全長転写とエキソン予測ゲノムの位置とEST構成データを組合わせてエキソンおよびその結果生じる転写が予測される。

【０１２１】

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

【０１２２】

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

【０１２３】

本発明はまた、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。

【０１２４】

本発明はまた、ＩＧＳＦＰをコードするポリヌクレオチドを含む。或る例では、本発明はＩＧＳＦＰをコードするＳＥＱ ＮＯ ＩＤ：７－１２からなる群から選択された配列を有するポリヌクレオチド配列を含む。ＳＥＱ ＮＯ ＩＤ：７－１２のポリヌクレオチド配列は、配列表に示すように等価ＲＮＡ配列を有しているが、窒素塩基チミンの出現はウラシルに置換され、糖のバックボーンはデオキシリボースではなくリボースから構成されている。

【０１２５】

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も含む。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：７－１２からなる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：７－１２からなる群から選択された配列を含むポリヌクレオチド配列の変異配列を提供する。 上記の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。

【０１２６】

前述の変異配列に加えて、あるいはその代わりとして、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に有意の配列同一性を有する部分を有し得るが、ｍＲＮＡ プロセッシング時にエキソンの選択的スプライシングから生じた配列ブロックの付加または欠失によりポリヌクレオチドの数はより多くなるか、またはより少なくなる。スプライス変異体は、その全長にわたって、ＩＧＳＦＰ をコードするポリヌクレオチド配列に対して約７０％より少ないか、あるいは約６０％より少ないか、または約５０％より少ないポリヌクレオチド配列同一性を有し得るが、スプライス変異体のいくつかの部分は、ＩＧＳＦＰをコードするポリヌクレオチド配列の部分に対して少なくとも約７０％、または、少なくとも約８５％、あるいは、約９５％ 、または１００％のポリヌクレオチド配列同一性を有する。上記の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つを有するアミノ酸配列をコードすることができる。

【０１２７】

遺伝暗号の縮重により作り出され得るＩＧＳＦＰをコードする多くのポリヌクレオチド配列には、自然発生する任意の既知の遺伝子のポリヌクレオチド配列と最小の類似性しか有しないものも含まれることを、当業者は理解するであろう。したがって本発明は、可能コドン選択に基づく組合せの選択によって産出し得るようなありとあらゆる可能性のあるポリヌクレオチド配列変異体を網羅し得る。これらの組み合わせは、天然の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、全ての変

異が明確に開示されているとみなす。

【０１２８】

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有する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こ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、特定の真核又は原核宿主に発生するペプチドの発現率を高めるようにコドンを選択することが可能である。コードされたアミノ酸配列を変えずに、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変更する別の理由は、天然の配列から作られる転写物より、例えば長い半減期など好ましい特性を備えるＲＮＡ転写物を作ることにある。

【０１２９】

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、当分野でよく知られている試薬を用いて、この合成配列を任意の様々な入手可能な発現ベクター及び細胞系中に挿入し得る。 更に、合成化学を用いてＩＧＳＦＰ またはその任意の断片をコードする配列に突然変異を誘導し得る。

【０１３０】

更に本発明には、種々のストリンジェントな条件下で、請求項に記載されたポリヌクレオチド配列、特に、ＳＥＱ ＩＤ ＮＯ：７－１２ 及びそれらの断片とハイブリダイズ可能なポリヌクレオチド配列が含まれる（例えば、Ｗａｈｌ，Ｇ．Ｍ．及びＳ．Ｌ．Ｂｅｒｇｅｒ（１９８７）Ｍｅｔｈｏｄｓ Ｅｎｚｙｍｏｌ．１５２：３９９－４０７、Ｋｉｍｍｅｌ，Ａ．Ｒ．（１９８７）Ｍｅｔｈｏｄｓ Ｅｎｚｙｍｏｌ．１５２：５０７－５１１等を参照）。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。

【０１３１】

ＤＮＡシークエンシングの方法は当分野でよく知られており、本発明の何れの実施例もＤＮＡシークエンシング方法を用いて実施可能である。ＤＮＡシークエンシング方法には酵素を用いることができ、例えばＤＮＡポリメラーゼＩのクレノウ断片、ＳＥＱＵＥＮＡＳＥ（ＵＳ Ｂｉｏｃｈｅｍｉｃａｌ，Ｃｌｅｖｅｌａｎｄ ＯＨ）、Ｔａｑポリメラーゼ（Ａｐｐｌｉｅｄ Ｂｉｏｓｙｓｔｅｍｓ）、熱安定性Ｔ７ポリメラーゼ（Ａｍｅｒｓｈａｍ，Ｐｈａｒｍａｃｉａ Ｂｉｏｔｅｃｈ，Ｐｉｓｃａｔａｗａｙ ＮＪ）を用いることができる。 或いは、例えばＥＬＯＮＧＡＳＥ増幅システム（Ｌｉｆｅ Ｔｅｃｈｎｏｌｏｇｉｅｓ，Ｇａｉｔｈｅｒｓｂｕｒｇ ＭＤ）において見られるように、ポリメラーゼと校正エキソヌクレアーゼを併用することができる。好しくは、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調製を自動化する。次に、ＡＢＩ ３７３ 或いは ３７７ ＤＮＡシークエンシングシステム（Ａｐｐｌｉｅｄ Ｂｉｏｓｙｓｔｅｍｓ）、ＭＥＧＡＢＡＣＥ １０００ ＤＮＡシークエンシングシステム（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クエンシングを行う。 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）。

【０１３２】

当分野で周知のＰＣＲ法をベースにした種々の方法で、部分的なヌクレオチド配列を利用

して、ＩＧＳＦＰ をコードする核酸配列を伸長し、プロモーターや調節エレメントなどの上流にある配列を検出する。例えば、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．２：３１８３２２を参照。）別の方法に逆ＰＣＲ法があり、これは広範な方向に伸長させたプライマーを用いて環状化した鋳型から未知の配列を増幅する方法である。鋳型は、既知のゲノム遺伝子座及びその周辺の配列を含む制限酵素断片から得る（例えば、Ｔｒｉｇｌｉａ，Ｔ．他（１９８８）Ｎｕｃｌｅｉｃ Ａｃｉｄｓ Ｒｅｓ．１６：８１８６参照）。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増幅する方法に関与している。（Ｌａｇｅｒｓｔｒｏｍ，Ｍ．他（１９９１）ＰＣＲ Ｍｅｔｈｏｄｓ Ａｐｐｌｉｃ １：１１１－１１９等を参照）。この方法では、ＰＣＲを行う前に複数の制限酵素の消化及び連結反応を用いて未知の配列領域内に組換え二本鎖配列を挿入することが可能である。また、未知の配列を検索するために用い得る他の方法については当分野で知られている。（Ｐａｒｋｅｒ，Ｊ．Ｄ．他（１９９１）Ｎｕｃｌｅｉｃ Ａｃｉｄｓ Ｒｅｓ．１９：３０５５－３０６０等を参照）。更に、ＰＣＲ、ネステッドプライマー及び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。

【０１３３】

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

【０１３４】

市販されているキャピラリー電気泳動システムを用いて、シークエンシングまたはＰＣＲ産物のサイズを分析し、またはそのヌクレオチド配列を確認することができる。具体的には、キャピラリーシークエンシングは、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、レーザで活性化される蛍光色素と、放出された波長の検出に利用するＣＣＤカメラとを有し得る。出力／光の強度は、適切なソフトウェア（Ａｐｐｌｉｅｄ Ｂｉｏｓｙｓｔｅｍｓ社のＧＥＮＯＴＹＰＥＲ、ＳＥＱＵＥＮＣＥ ＮＡＶＩＧＡＴＯＲ等）を用いて電気信号に変換し得る。 サンプルのロードからコンピュータ分析及び電子データ表示までの全プロセスがコンピュータ制御可能である。キャピラリー電気泳動法は、特定のサンプルに少量しか存在しないようなＤＮＡ小断片のシークエンシングに特に適している。

【０１３５】

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配列またはその断片を組換えＤＮＡ分子にクローニングして、適切な宿主細胞内にＩＧＳＦＰ、その断片または機能的等価物を発現させることが可能である。遺伝暗号固有の縮重により、実質的に同じ或いは機能的に等価のアミノ酸配列をコードする別のＤＮＡ配列を作ることができ、これらの配列をＩＧＳＦＰの発現に利用可能である。

【０１３６】

種々の目的でＩＧＳＦＰをコードする配列を変えるために、当分野で一般的に知られてい

る方法を用いて、本発明のヌクレオチド配列を組換えることができる。 この目的には、遺伝子産物のクローン化、プロセッシング及び／または発現の調節が含まれるが、これらに限定されるものではない。遺伝子断片及び合成オリゴヌクレオチドのランダムなフラグメンテーション及びＰＣＲ再アセンブリによるＤＮＡシャッフリングを用い、ヌクレオチド配列を組み換えることが可能である。例えば、オリゴヌクレオチド仲介特定部位突然変異誘導を利用して、新規な制限部位の生成、グリコシル化パターンの変更、コドン優先の変更、スプライス変異体の生成等を起こす突然変異を導入し得る。

【０１３７】

本発明のヌクレオチドを、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；Ｃｒａｍｅｒｉ，Ａ．他（１９９６）Ｎａｔ．Ｂｉｏｔｅｃｈｎｏｌ．１４：３１５－３１９）などのＤＮＡシャフリング技術を用いてシャフリングして、ＩＧＳＦＰの生物学的または酵素的な活性、或いは他の分子や化合物と結合する能力などのＩＧＳＦＰの生物学的特性を変更或いは改良することができる。ＤＮＡシャッフリングは、遺伝子断片のＰＣＲ仲介再組換えを用いて遺伝子変異体のライブラリを生成するプロセスである。ライブラリはその後、その遺伝子変異体を所望の特性に同定するような選択またはスクリーニングにかける。次にこれらの好適な変異体をプールし、更に反復してＤＮＡシャッフリング及び選択／スクリーニングを行ってもよい。このように、遺伝の多様性は「人為的」品種改良及び急速な分子の進化を経て創生される。例えば、ランダムポイント突然変異を有する単一の遺伝子の断片を組み換えて、スクリーニングし、その後所望の特性が最適化されるまでシャッフリングすることができる。或いは、所定の遺伝子を同種または異種のいずれかから得た同一遺伝子ファミリーの相同遺伝子と組み換えて、それによって天然に存在する複数の遺伝子の遺伝多様性を、指図された制御可能な方法で最大化させることができる。

【０１３８】

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；及び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 ２６９：２０２２０４等を参照）。自動合成はＡＢＩ ４３１Ａペプチドシンセサイザ（Ｐｅｒｋｉｎ Ｅｌｍｅｒ）を用いて達成し得る。更にＩＧＳＦＰのアミノ酸配列または任意のその一部は、直接的な合成の際の変更、及び／または化学的方法を用いた他のタンパク質または任意のその一部からの配列との組み合わせにより、天然のポリペプチド配列を有するポリペプチドまたは変異体ポリペプチドを作製することが可能である。

【０１３９】

ペプチドは、分離用高速液体クロマトグラフィーを用いて実質上精製可能である（Ｃｈｉｅｚ，Ｒ．Ｍ．および 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クエ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）。

【０１４０】

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。このような要素は、長さ及び特異性が様々である。特定の開始シグナルによって、ＩＧＳＦＰをコードする配列のより効果的な翻訳を達成することが可能である。このようなシグナルには、ＡＴＧ開始コドンと、コザック配列などの近傍の配列が含まれる。ＩＧＳＦＰ をコードする配列及びその開始コドン、上流の調節配列が好適な発現ベクターに挿入された場合は、更なる転写調節シグナルや翻訳調節シグナルは必要なくなるであろう。しかしながら、コーディング配列或いはその断片のみが挿入された場合は、インフレームのＡＴＧ開始コドンを含む外来性の翻訳調節シグナルが発現ベクターに含まれるようにすべきである。外来性の翻訳要素及び開始コドンは、様々な天然物及び合成物を起源とし得る。用いられる特定の宿主細胞系に好適なエンハンサーを含めることで発現の効率を高めることが可能である。（例えば、Ｓｃｈａｒｆ，Ｄ．他（１９９４）Ｒｅｓｕｌｔｓ Ｐｒｏｂｌ．Ｃｅｌｌ Ｄｉｆｆｅｒ．２０１－１８－１６２．を参照）。

【０１４１】

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章及び８章，及び１６－１７章；及びＡｕｓｕｂｅｌ，Ｆ．Ｍ．他．（１９９５）Ｃｕｒｒｅｎｔ Ｐｒｏｔｏｃｏｌｓ ｉｎ Ｍｏｌｅｃｕｌａｒ Ｂｉｏｌｏｇｙ，Ｊｏｈｎ Ｗｉｌｅｙ ＆ Ｓｏｎｓ，Ｎｅｗ Ｙｏｒｋ ＮＹ，９章、１３章及び１６章を参照）。

【０１４２】

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や、酵母菌発現ベクターで形質転換させた酵母菌や、ウイルス発現ベクター（例えばバキュロウイルス）に感染した昆虫細胞系や、ウイルス発現ベクター（例えばカリフラワーモザイクウイルスＣａＭＶまたはタバコモザイクウイルスＴＭＶ）または細菌発現ベクター（例えばＴｉまたはｐＢＲ３２２プラスミド）で形質転換させた植物細胞系、動物細胞系などの微生物等がある。（前出のＳａｍｂｒｏｏｋ、前出のＡｕｓｕｂｅｌ、Ｖａｎ Ｈｅｅｋｅ，Ｇ．ａｎｄ Ｓ．Ｍ．Ｓｃｈｕｓｔｅｒ（１９８９）Ｊ．Ｂｉｏｌ．Ｃｈｅｍ．２６４：５５０３－５５０９、；Ｅｎｇｅｌｈａｒｄ、Ｅ．Ｋ．他（１９９４）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ ９１：３２２４－３２２７、Ｓａｎｄｉｇ，Ｖ．他（１９９６）Ｈｕｍ．Ｇｅｎｅ Ｔｈｅｒ．７：１９３７－１９４５、Ｔａｋａｍａｔｓｕ，Ｎ．（１９８７）ＥＭＢＯＪ．６：３０７－３１１、；『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．ａｎｄ 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輸送することができる（Ｄｉ Ｎｉｃｏｌａ，Ｍ．他（１９９８）Ｃａｎｃｅｒ Ｇｅｎ．Ｔｈｅｒ．５（６）：３５０－３５６、Ｙｕ，Ｍ．他（１９９３）Ｐｒｏｃ．Ｎ

ａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ９０（１３）：６３４０－６３４４、Ｂｕｌｌｅｒ，Ｒ．Ｍ．他（１９８５）Ｎａｔｕｒｅ３１７（６０４０）：８１３－８１５；ＭｃＧｒｅｇｏｒ，Ｄ．Ｐ．他（１９９４）Ｍｏｌ．Ｉｍｍｕｎｏｌ．３１（３）：２１９－２２６、Ｖｅｒｍａ，Ｉ．Ｍ．及びＮ．Ｓｏｍｉａ（１９９７）Ｎａｔｕｒｅ３８９：２３９－２４２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。

【０１４３】

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の日常的なクローニング、サブクローニング、増殖には、ＰＢＬＵＥＳＣＲＩＰＴ（Ｓｔｒａｔａｇｅｎｅ，Ｌａ Ｊｏｌｌａ ＣＡ）またはｐＳＰＯＲＴ１プラスミド（ＧＩＢＣＯ ＢＲＬ）などの多機能の大腸菌ベクターを用いることができる。ベクターの多数のクローニング部位にＩＧＳＦＰ をコードする配列をライゲーションするとｌａｃＺ遺伝子が破壊され、組換え分子を含む形質転換された細菌の同定のための比色スクリーニング法が可能となる。更にこれらのベクターは、クローニングされた配列におけるｉｎ ｖｉｔｒｏ転写、ジデオキシのシークエンシング、ヘルパーファージによる一本鎖の救出、入れ子状態の欠失の生成にも有用であろう（例えば、Ｖａｎ Ｈｅｅｋｅ，Ｇ．およびＳ．Ｍ．Ｓｃｈｕｓｔｅｒ（１９８９）Ｊ．Ｂｉｏｌ．Ｃｈｅｍ．２６４：５５０３５５０９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 の発現をハイレベルで誘導するベクターが使用できる。例えば、強力な誘導ＳＰ６バクテリオファージプロモーターまたは誘導Ｔ７バクテリオファージプロモーターを含むベクターが使用できる。

【０１４４】

ＩＧＳＦＰの産生に酵母の発現系の使用が可能である。α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（Ｓａｃｃｈａｒｏｍｙｃｅｓ ｃｅｒｅｖｉｓｉａｅ）またはピキア酵母（Ｐｉｃｈｉａ ｐａｓｔｏｒｉｓ）に使用可能である。更に、このようなベクターは、発現したタンパク質の分泌か細胞内への保持のどちらかを誘導し、安定した増殖のために宿主ゲノムの中に外来配列を組み込む。（例えば、Ａｕｓｕｂｅｌ，１９９５，前出、Ｂｉｔｔｅｒ，Ｇ．Ａ．他（１９８７）Ｍｅｔｈｏｄｓ Ｅｎｚｙｍｏｌ．１５３：５１６－５４４、及びＳｃｏｒｅｒ．Ｃ．Ａ．他（１９９４）Ｂｉｏ／Ｔｅｃｈｎｏｌｏｇｙ １２１－１８１－１８４．を参照）。

【０１４５】

植物系もＩＧＳＦＰの発現に使用可能である。ＩＧＳＦＰ をコードする配列の転写は、ウイルスプロモーター、例えば単独あるいはＴＭＶ由来のオメガリーダー配列と組み合せて用いられるようなＣａＭＶ由来の３５Ｓおよび１９Ｓプロモーターによって促進される（Ｔａｋａｍａｔｓｕ，Ｎ．（１９８７）ＥＭＢＯ Ｊ．６：３０７３１１）。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．３：１６７１１６８０、Ｂｒｏｇｌｉｅ，Ｒ．他（１９８４）Ｓｃｉｅｎｃｅ ２２４：８３８８４３、Ｗｉｎｔｅｒ，Ｊ．他（１９９１）Ｒｅｓｕｌｔｓ Ｐｒｏｂｌ．Ｃｅｌｌ Ｄｉｆｆｅｒ．１７：８５１０５等を参照）。これらの構成物は、直接ＤＮＡ形質転換または病原体を媒介とする形質移入によって、植物細胞内に導入可能である（『マグローヒル科学技術年鑑』（Ｔｈｅ ＭｃＧｒａｗ Ｈｉｌｌ Ｙｅａｒｂｏｏｋ ｏｆ Ｓｃｉｅｎｃｅ ａｎｄ Ｔｅｃｈｎｏｌｏｇｙ）（１９９２）ＭｃＧｒａｗ Ｈｉｌｌ Ｎｅｗ Ｙｏｒｋ ＮＹ，１９１－１９６ページ等を参照）。

【０１４６】

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結合し得る。非必

須のＥ１またはＥ３領域へウイルスのゲノムを挿入し、宿主細胞で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．およびＴ．Ｓｈｅｎｋ（１９８４）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ８１：３６５５３６５９を参照）。更に、ラウス肉腫ウイルス（ＲＳＶ）エンハンサー等の転写エンハンサーを用いて、哺乳動物宿主細胞における発現を増大させ得る。ＳＶ４０またはＥＢＶをベースにしたベクターを用いてタンパク質を高レベルで発現させることもできる。

【０１４７】

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

【０１４８】

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、発現ベクターを用いて、ＩＧＳＦＰ をコードする配列を株化細胞に形質転換することが可能である。このような発現ベクターは、ウイルス起源の複製及び／または内在性の発現要素や、同じ或いは別のベクターの上の選択マーカー遺伝子を含む。ベクターの導入後、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。

【０１４９】

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$^{-}$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単純ヘルペスウイルスのアデニンホスホリボシルトランスフェラーゼ遺伝子がある（例えば、Ｗｉｇｌｅｒ，Ｍ．他（１９７７）Ｃｅｌｌ １１：２２３－２３２；及びＬｏｗｙ，Ｉ．他（１９８０）Ｃｅｌｌ ２２：８１７－８２３を参照）。また、選択の基礎として代謝拮抗物質、抗生物質或いは除草剤への耐性を用いることができる。例えばｄｈｆｒはメトトレキセートに対する耐性を与え、ｎｅｏはアミノグリコシッドネオマイシン及びＧ－４１８に対する耐性を与え、ａｌｓはクロルスルフロンに対する耐性を、ｐａｔはホスフィノトリシンアセチルトランスフェラーゼに対する耐性を各々与える（Ｗｉｇｌｅｒ，Ｍ．他（１９８０）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ７７：３５６７３５７０；ＣｏｌｂｅｒｅＧａｒａｐｉｎ，Ｆ．他（１９８１）Ｊ．Ｍｏｌ．Ｂｉｏｌ．１５０：１１４等を参照。）この他の選択可能な遺伝子、例えば、代謝のための細胞の必要条件を変えるｔｒｐＢ及びｈｉｓＤは、文献に記載されている（例えば、Ｈａｒｔｍａｎ，Ｓ．Ｃ．及びＲ．Ｃ．Ｍｕｌｌｉｇａｎ（１９８８）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ８５：８０４７８０５１参照。）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）、βグルクロニダーゼ及びその基質βグルクロニド、またはルシフェラーゼ及びその基質ルシフェリン等を用いてもよい。これらのマーカーを用いて、トランスフォーマントを特定するだけでなく、特定のベクター系に起因する一過性或いは安定したタンパク質発現を定量することが可能である（Ｒｈｏｄｅｓ，Ｃ．Ａ．（１９９５）Ｍｅｔｈｏｄｓ Ｍｏｌ．Ｂｉｏｌ．５５：１２１１３１等を参照）。

【０１５０】

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、１つのプ

ロモーターの制御下でマーカー遺伝子がＩＧＳＦＰ をコードする配列とタンデムに配置することも可能である。誘導または選択に応答したマーカー遺伝子の発現は通常、タンデム遺伝子の発現も示す。

【０１５１】

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

【０１５２】

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、当分野で周知である。 このような技法には、酵素に結合した免疫吸着剤検定法（ＥＬＩＳＡ）、ラジオイムノアッセイ（ＲＩＡ）、フローサイトメーター（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，Ｈｕｍａｎｓ Ｐｒｅｓｓ，Ｔｏｔｏｗａ ＮＪ等を参照）。

【０１５３】

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

【０１５４】

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。

【０１５５】

更に、宿主細胞株の選択は、挿入した配列の発現を調節する能力または発現したタンパク

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

【０１５６】

本発明の別の実施例では、ＩＧＳＦＰ をコードする自然或いは変更された、または組換えの核酸配列を上記した任意の宿主系の融合タンパク質の翻訳となる異種配列に結合させる。例えば、市販の抗体によって認識され得る異種部分を含むキメラＩＧＳＦＰ タンパク質が、ＩＧＳＦＰ の活性のインヒビターに対するペプチドライブラリのスクリーニングを促進し得る。また、異種タンパク質部分及び異種ペプチド部分は、また市販の親和性基質を用いた融合タンパク質の精製を促進する。限定されるものではないがこのような部分には、グルタチオンＳトランスフェラーゼ（ＧＳＴ）、マルトース結合タンパク質（ＭＢＰ）、チオレドキシン（Ｔｒｘ）、カルモジュリン結合ペプチド（ＣＢＰ）、６－Ｈｉｓ、ＦＬＡＧ、ｃ－ｍｙｃ、赤血球凝集素（ＨＡ）がある。ＧＳＴは固定化グルタチオン上で、ＭＢＰはマルトース上で、Ｔｒｘはフェニルアルシンオキシド上で、ＣＢＰはカルモジュリン上で、そして６－Ｈｉｓは金属キレート樹脂上で、同族の融合タンパク質の精製を可能にする。ＦＬＡＧ、ｃ－ｍｙｃ及び赤血球凝集素（ＨＡ）は、これらのエピトープ標識を特異的に認識する市販のモノクローナル抗体及びポリクローナル抗体を用いて、融合タンパク質の免疫親和性精製を可能にする。また、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。

【０１５７】

本発明の別の実施例では、ＩＧＳＦＰ 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 の合成が行われる。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結合する。翻訳は、例えば^{35}Ｓメチオニンのような放射能標識したアミノ酸前駆体の存在下で起こる。

【０１５８】

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。

【０１５９】

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、化合物は、ＩＧＳＦＰ が結合する天然受容体、或いは例えばリガンド結合部位などの少なくとも受容体のある断片に密接に関連し得る。何れの場合も、既知の技術を用いてこの化合物を合理的に設計することができる。一実施例では、このような化合物に対するスクリーニングには、分泌タンパク質或いは細胞膜上のタンパク質の何れか一方として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、大腸菌からの細胞が含まれる。ＩＧＳＦＰ を発現する細胞またはＩＧＳＦＰ を含有する細胞膜断片を試験化合物と接触させて、ＩＧＳＦＰ または化合物の何れかの結合、刺激または阻害を分析する。

【０１６０】

あるアッセイは、単に試験化合物をポリペプチドに実験的に結合させ、結合を、蛍光色素、放射性同位体、酵素抱合体またはその他の検出可能な標識により検出することができる。例えば、このアッセイは、少なくとも１つの試験化合物を、溶液中の或いは固体支持物に固定されたＩＧＳＦＰ と結合させるステップと、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、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。

【０１６１】

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することが可能である。このような化合物には、アゴニスト、アンタゴニスト、或るいは部分的または逆アゴニスト等が含まれる。一実施例では、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。別法では、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。これらアッセイの何れかにおいて、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つから複数の試験化合物をスクリーニングすることができる。

【０１６２】

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

【０１６３】

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、中胚葉及び外胚葉の細胞の種類を含む少なくとも８つの別々の細胞系統に分化する可能性を有する。これらの細胞系統は、例えば神経細胞、造血系統及び心筋細胞に分化する（Ｔｈｏｍｓｏｎ，Ｊ．Ａ．他（１９９８）Ｓｃｉｅｎｃｅ ２８２：１１４５－１１４７）。

【０１６４】

また、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）または遺伝子組み換え動物（マウスまたはラット）を作製することも可能である。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研究し、可能性のある医薬品を用いて処理し、ヒトの疾患の治療に関する情報を得る。別法では、例えば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）。

【０１６５】

（治療）

ＩＧＳＦＰ の領域と免疫グロブリンスーパファミリタンパク質領域との間に、例えば配列及びモチーフのからみで化学的及び構造的類似性が存在する。さらに、ＩＧＳＦＰの発現は好酸球、肺組織、前立腺腫瘍組織、子宮筋組織および淡蒼球や無名質のような脳組織に密接に関連している。したがって、ＩＧＳＦＰは免疫系疾患、神経疾患、発達障害、筋肉疾患および癌を含む細胞増殖異常において役割を果たしていると考えられる。ＩＧＳＦＰ の発現若しくは活性の増大に関連する疾患の治療においては、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においては、ＩＧＳＦＰの発現または活性を増大させることが望ましい。

【０１６６】

従って、或る実施例において、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、患者にＩＧＳＦＰ またはその断片や誘導体を投与することが可能である。そのような疾患の例として、限定するものではないが、免疫疾患の中には、後天性免疫不全症候群（ＡＩＤＳ）、ＢｒｕｔｏｎのＸ染色体連関無ガンマグロブリン血症、分類不能型免疫不全（ＣＶＩ）、デジョージ症候群（胸腺発育不全）、胸腺形成異常、分離型ＩｇＡ欠損症、重症複合型免疫不全疾患（ＳＣＩＤ）、血小板減少症および湿疹を伴う免疫不全（ヴィスコット‐オールドリッチ症候群、Ｃｈｅ´ｄｉａｋ‐東症候群、慢性肉芽腫症、遺伝性血管〔運動〕神経性水腫（浮腫）、クッシング病合併免疫不全症、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、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫性多腺性内分泌カンジダ性外胚葉ジストロフィー（ＡＰＥＣＥＤ）、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、糖尿病、肺気腫、リンパ球毒素性一時性リンパ球減少症、赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性大腸症候群、多発性硬化症、重症筋無力症、心筋または心膜炎症、骨関節炎、骨粗しょう症、膵炎、多発性筋炎、乾癬、ライター症候群、リウマチ様関節炎、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、血小板減少性紫斑病、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌合併症、血液透析、体外循環、ウイルス感染症、細菌感染症、真菌感染症、寄生虫感染症、原虫感染症、蠕虫感染症、外傷が含まれ、神経系疾患の中には、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、パーキソン病及びその他の錐体外路障害、筋萎縮性側策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性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、及びトゥーレット病と、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が含まれ、また発達障害も含まれ、そ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、Ｓｙｎｄｅｎｈａｍ舞踏病（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、筋肉疾患には、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、筋緊張性ジストロフィー、中核疾患、ネマリンミオパシー、橋中心髄鞘崩壊症、脂質筋疾患、ミトコンドリアミオパシー、感染性筋炎、多発性筋炎、皮膚性筋炎、封入体筋炎、甲状腺中毒筋炎およびエタノールミオパシーが含まれ、また、細胞増殖異常では日光性角化症及びアテローム性動脈硬化、滑液包炎、硬変、肝炎、混合型結合組織病（ＭＣＴＤ）、骨髄線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症、並びに腺癌及び白血病、リンパ腫、黒色腫、骨髄腫、肉腫、及び奇形癌、具体的には、副腎、膀胱、骨、骨髄、脳、乳房、頚部、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。

【０１６７】

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、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することも可能である。

【０１６８】

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。

【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、ＩＧＳＦＰ の活性を調節するアゴニストを患者に投与することも可能である。

【０１７０】

更なる実施例では、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、患者にＩＧＳＦＰ のアンタゴニストを投与することが可能である。このような疾患の例には、限定するものではないが、上記の免疫系疾患、神経系疾患、発達異常、筋疾患および癌を含む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ッティング或いは送達機構として間接的に用いられ得る。

【０１７１】

別の実施例では、上に列記した疾患に限らず、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患者に投与することも可能である。

【０１７２】

別の実施例では、本発明の任意のタンパク質、アンタゴニスト、抗体、アゴニスト、相補配列、またはベクターを、別の好適な治療薬と組み合わせて投与することもできる。併用療法で用いる好適な治療薬は、当業者が従来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。この方法を用いることにより少量の各薬剤で医薬効果をあげることが可能となり、それによって副作用の可能性を低減し得る。

【０１７３】

ＩＧＳＦＰ のアンタゴニストは、当分野で一般的な方法を用いて製造することが可能である。 詳しくは、精製されたＩＧＳＦＰ を用いて抗体を作ったり、治療薬のライブラリをスクリーニングしてＩＧＳＦＰ と特異的に結合するものの同定が可能である。ＩＧＳＦＰ の抗体も、当分野で一般的な方法を用いて製造することが可能である。 このような抗体には、ポリクローナル抗体、モノクローナル抗体、キメラ抗体、一本鎖、Ｆａｂフラグメント、及びＦａｂ発現ライブラリによって作られたフラグメントが含まれる。 但し、これらに限定されるものではない。中和抗体（即ち二量体の形成を阻害する抗体）は通常、治療用に好適である。

【０１７４】

抗体の産生のためには、ヤギ、ウサギ、ラット、マウス、ヒト及びその他のものを含む種々の宿主が、ＩＧＳＦＰ または任意の断片、または免疫原性の特性を備えるそのオリゴペプチドの注入によって免疫化され得る。宿主の種に応じて、種々のアジュバントを用いて免疫応答を高めることもできる。限定するものではないがこのようなアジュバントには、フロイントアジュバントと、水酸化アルミニウム等のミネラルゲルアジュバントと、リゾレシチン、プルロニックポリオル、ポリアニオン、ペプチド、油性乳剤、スカシガイのヘモシニアン、ジニトロフェノール等の界面活性剤とがある。ヒトに用いられるアジュバントの中では、ＢＣＧ（カルメット‐ゲラン杆菌）及びコリネバクテリウム‐パルヴムが特に好ましい。

【０１７５】

ＩＧＳＦＰ 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。また、これらのオリゴペプチド、ペプチドまたは断片は、天然のタンパク質のアミノ酸配列の一部と同一であることが好ましい。ＩＧＳＦＰ のアミノ酸の短いストレッチは、ＫＬＨなどの別のタンパク質の配列と融合し、キメラ分子に対する抗体が産生され得る。

【０１７６】

ＩＧＳＦＰ 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」作製のために発達した技術が、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（例えば、Ｍｏｒｒｉｓｏｎ，Ｓ．Ｌ．他．（１９８４）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．８１：６８５１６８５５；Ｎｅｕｂｅｒｇｅｒ，Ｍ．Ｓ．他．（１９８４）Ｎａｔｕｒｅ ３１２：６０４－６０８；Ｔａｋｅｄａ，Ｓ．他．（１９８５）Ｎａｔｕｒｅ ３１４：４５２－４５４を参照）。別法では、当分野で周知の方法を用いて、一本鎖抗体の産生のための記載された技術を適用して、ＩＧＳＦＰ特異性一本鎖抗体を生成する。関連特異性を有するがイディオタイプ組成が異なるような抗体を、ランダムな組合せの免疫グロブリンライブラリからチェーンシャッフリングによって産生することもできる（Ｂｕｒｔｏｎ Ｄ．Ｒ．（１９９１）Ｐｒｏｃ．Ｎａｔ

ｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ８８：１０１３４－１０１３７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、或いは免疫グロブリンライブラリのスクリーニングまたは文献に開示されているような高特異結合試薬のパネルのスクリーニングによっても行い得る。（Ｏｒｌａｎｄｉ，Ｒ．他（１９８９）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ８６：３８３３－３８３７、Ｗｉｎｔｅｒ，Ｇ．他（１９９１）Ｎａｔｕｒｅ ３４９：２９３－２９９等を参照）。

【０１７９】

ＩＧＳＦＰに対する特異的な結合部位を含む抗体も得ることができる。例えば、このような断片には、限定するものではないが、抗体分子のペプシン消化によって作製されるＦ（ａｂ’）$_2$断片と、Ｆ（ａｂ’）$_2$断片のジスルフィド架橋を還元することによって作製されるＦａｂ断片とがある。或いは、Ｆａｂ発現ライブラリを作製することによって、モノクローナルＦａｂ断片を所望の特異性で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。隔離された特異性を有するポリクローナル抗体またはモノクローナル抗体の何れかを用いる競合的な結合、または免疫放射線活性のための多くのプロトコルが、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、ＩＧＳＦＰ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。二つの非干渉性ＩＧＳＦＰエピトープに反応性のモノクローナル抗体を用いる、２部位モノクローナルベースのイムノアッセイが一般に利用されるが、競合的結合アッセイも利用することができる（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、ＩＧＳＦＰに対する抗体の親和性を評価する。親和性を結合定数Ｋａで表すが、このＫａは、平衡状態の下でＩＧＳＦＰ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。複数の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Ｋａは、ＩＧＳＦＰに対する抗体の平均親和性または結合活性を表す。特定のＩＧＳＦＰエピトープに単一特異的なモノクローナル抗体試薬のＫａは、親和性の真の測定値を表す。Ｋａ値が10^9～10^{12}ｌｉｔｅｒ／ｍｏｌの高親和性抗体試薬は、ＩＧＳＦＰ抗体複合体が激しい操作に耐えなければならないイムノアッセイに用いるのが好ましい。Ｋａ値が10^6～10^7ｌｉｔｅｒ／ｍｏｌの低親和性抗体試薬は、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及び類似の処理に用いるのが好ましい（Ｃａｔｔｙ，Ｄ．（１９８８）Ａｎｔｉｂｏｄｉｅｓ，Ｖｏｌｕｍｅ Ｉ：Ａ Ｐｒａｃｔｉｃａｌ Ａｐｐｒｏａｃｈ．ＩＲＬ Ｐｒｅｓｓ，Ｗａｓｈｉｎｇｔｏｎ，ＤＣ；Ｌｉｄｄｅｌｌ，Ｊ．Ｅ．及びＣｒｙｅｒ，Ａ．（１９９１）Ａ Ｐｒａｃｔｉｃａｌ Ｇｕｉｄｅ ｔｏ Ｍｏｎｏｃｌｏｎａｌ Ａｎｔｉｂｏｄｉｅｓ，Ｊｏｈｎ Ｗｉｌｅｙ ＆ Ｓｏｎｓ，Ｎｅｗ Ｙｏｒｋ ＮＹ）。

【０１８２】

ポリクローナル抗体試薬の抗体価及び結合活性を更に評価して、後に使う或る適用例に対するこのような試薬の品質及び適性を決定することができる。例えば、少なくとも１～２ｍｇ／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医薬は一般に、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、同Ｃｏｌｉｇａｎ他を参照）。

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

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

【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染色体連関遺伝（Ｃａｖａｚｚａｎａ－Ｃａｌｖｏ，Ｍ．他（２０００）Ｓｃｉｅｎｃｅ ２８８：６６９－６７２）により特徴付けられる重度の複合型免疫欠損症（ＳＣＩＤ）－Ｘ１の場合）、先天性アデノシンデアミナーゼ（ＡＤＡ）欠損症に関連する重度の複合型免疫欠損症（Ｂｌａｅｓｅ，Ｒ．Ｍ．他（１９９５）Ｓｃｉｅｎｃｅ ２７０：４７５－４８０、Ｂｏｒｄｉｇｎｏｎ，Ｃ．他（１９９５）Ｓｃｉｅｎｃｅ ２７０：４７０－４７５）、嚢胞性繊維症（Ｚａｂｎｅｒ，Ｊ．他（１９９３）Ｃｅｌｌ ７５：２０７－２１６：Ｃｒｙｓｔａｌ、Ｒ．Ｇ．他（１９９５）Ｈｕｍ．Ｇｅｎｅ Ｔｈｅｒａｐｙ ６：６４３－６６６、Ｃｒｙｓｔａｌ，Ｒ．Ｇ．他．（１９９５）Ｈｕｍ．Ｇｅｎｅ Ｔｈｅｒａｐｙ ６：６６７－７０３）、サラセミア（ｔｈａｌａｓｓａｍｉａ）、家族性高コレステロール血症、および第ＶＩＩＩ因子若しくは第ＩＸ因子欠損に起因する血友病（Ｃｒｙｓｔａｌ，３５ Ｒ．Ｇ．（１９９５）Ｓｃｉｅｎｃｅ ２７０：４０４－４１０、Ｖｅｒｍａ，Ｉ．Ｍ．およびＳｏｍｉａ．Ｎ．（１９９７）Ｎａｔｕｒｅ ３８９：２３９－２４２）を治療し、（ｉｉ）条件的致死性遺伝子産物を発現させ（例えば制御不能な細胞増殖に起因する癌の場合）、（ｉｉｉ）細胞内の寄生生物（例えばヒト免疫不全ウイルス（ＨＩＶ）（Ｂａｌｔｉｍｏｒｅ，Ｄ．（１９８８）Ｎａｔｕｒｅ ３３５：３９５－３９６、Ｐｏｅｓｃｂｌａ，Ｅ．他（１９９６）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ．９３：１１３９５－１１３９９）、Ｂ型若しくはＣ型肝炎ウイルス（ＨＢＶ、ＨＣＶ）、Ｃａｎｄｉｄａ ａｌｂｉｃａｎｓ及びＰａｒａｃｏｃｃｉｄｉｏｉｄｅｓ ｂｒａｓｉｌｉｅｎｓｉｓ等の真菌寄生菌、並びに熱帯熱マラリア原虫及びクルーズトリパノソーマ等の原虫寄生生物に対する防御機能を有するタンパク質を発現させることができる。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臨床症状の発現を緩和することが可能である。

【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。

【０１８７】

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）、ＰＣＭＶ－ＳＣＲＩＰＴ、ＰＣＭＶ－ＴＡＧ、ＰＥＧＳＨ／ＰＥＲＶ（Ｓｔｒａｔａｇｅｎｅ，Ｌａ Ｊｏｌｌａ ＣＡ）、ＰＴＥＴ－ＯＦＦ、ＰＴＥＴ－ＯＮ、ＰＴＲＥ２、ＰＴＲＥ２－ＬＵＣ、ＰＴＫ－ＨＹＧ（Ｃｌｏｎｔｅｃｈ，Ｐａｌｏ Ａｌｔｏ ＣＡ）が含まれる。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β－アクチン遺伝子等）、（ｉｉ）誘導性プロモーター（例えば、市販されているＴ－ＲＥＸプラスミド（Ｉｎｖｉｔｒｏｇｅｎ）に含まれている、テトラサイクリン調節性プロモーター（Ｇｏｓｓｅｎ，Ｍ．ａｎｄ Ｈ．Ｂｕｊａｒｄ（１９９２）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．Ｕ．Ｓ．Ａ．８９：５５４７－５５５１；Ｇｏｓｓｅｎ，Ｍ．他（１９９５）Ｓｃｉｅｎｃｅ ２６８：１７６６－１７６９；Ｒｏｓｓｉ，Ｆ．Ｍ．Ｖ．ａｎｄ Ｈ．Ｍ．Ｂｌａｕ（１９９８）Ｃｕｒｒ．Ｏｐｉｎ．Ｂｉｏｔｅｃｈｎｏｌ．９：４５１－４５６））、エクジソン誘導性プロモーター（市販されているプラスミドＰＶＧＲＸＲ及びＰＩＮＤに含まれている：Ｉｎｖｉｔｒｏｇｅｎ）、ＦＫ５０６／ラパマイシン誘導性プロモーター、またはＲＵ４８６／ミフェプリストーン誘導性プロモーター（Ｒｏｓｓｉ，Ｆ．Ｍ．Ｖ．ａｎｄ Ｈ．Ｍ．Ｂｌａｕ，前出）、または（ｉｉｉ）正常な個体に由来するＩＧＳＦ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。

【０１８８】

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

【０１８９】

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ＰＦＢＮＥＯ）はＳｔｒａｔａｇｅｎｅ社から市販されており、刊行データ（Ｒｉｖｉｅｒｅ，Ｉ．他（１９９５）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．Ｕ．Ｓ．Ａ．９２：６７３３－６７３７）に基づいている。上記データを引用することをもって本明細書の一部とする。ベクターは、好適なベクター産生細胞系（ＶＰＣＬ）において増殖され、ＶＰＣＬは、標的細胞上の受容体に対する親和性を有するエンベロープ遺伝子またはＶＳＶｇ等の汎親和性エンベロープタンパク質を発現する（Ａｒｍｅｎｔａｎｏ，Ｄ．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系を得るための方法が開示されており、引用することをもって本明細書の一部とする。レトロウイルスベクターの増殖、細胞集団（例えばＣＤ４$^{+}$ Ｔ細胞）の形質導入、及び形質導入した細胞の患者への戻しは、遺伝子治療の分野では当業者に公知の方法であり、多数の文献に記載されている（Ｒａｎｇａ，Ｕ．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）。

【０１９０】

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つ或いは複数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の作製及びパッケージングについては、当業者に公知である。複製欠損型アデノウイルスベクターは、免疫調節タンパク質をコードする遺伝子を膵臓の無損傷の膵島内に導入するために可変性であることが証明された（Ｃｓｅｔｅ，Ｍ．Ｅ．他（１９９５）Ｔｒａｎｓｐｌａｎｔａｔｉｏｎ ２７：２６３－２６８）。使用できる可能性のあるアデノウイルスベクターは、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 及び Ｖｅｒｍａ，Ｉ．Ｍ．およびＮ．Ｓｏｍｉａ（１９９７）Ｎａｔｕｒｅ １８：３８９：２３９－２４２も参照されたい。両文献は、引用することをもって本明細書の一部とする。

【０１９１】

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つ或いは複数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）系のベクターのは、ＨＳＶ親和性の中枢神経細胞にＩＧＳＦＰ を導入する際に特に有用である。ヘルペス系ベクターの作製及びパッケージングは、当業者に公知である。複製適格性単純ヘルペスウイルス（ＨＳＶ）Ｉ型系のベクターは、レポーター遺伝子を霊長類の眼に送達するために用いられてきた（Ｌｉｕ，Ｘ．他（１９９９）Ｅｘｐ．Ｅｙｅ Ｒｅｓ．１６９：３８５－３９５）。ＨＳＶ－１ウイルスベクターの作製についても、ＤｅＬｕｃａに付与された米国特許第５，８０４，４１３号（”Ｈｅｒｐｅｓ ｓｉｍｐｌｅｘ ｖｉｒｕｓ ｓｗａｉｎｓ ｆｏｒ ｇｅ

ｎｅ 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つの外来性遺伝子を有するゲノムを含む組換えＨＳＶ ｄ９２についての記載がある。上記特許はまた、ＩＣＰ４、ＩＣＰ２７及びＩＣＰ２２のために除去される組換えＨＳＶ系統の作製及び使用について開示している。ＨＳＶベクターについては、Ｇｏｉｎｓ，Ｗ．Ｆ．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部分を含む多数のプラスミドを形質移入した後の組換えウイルスの産生、ヘルペスウイルスの成長及び増殖、並びにヘルペスウイルスの細胞への感染は、当業者に公知の技術である。

【０１９２】

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

【０１９３】

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

【０１９４】

リボザイムは酵素性ＲＮＡ分子であり、ＲＮＡの特異的切断を触媒するためにリボザイムを用いることもできる。リボザイム作用のメカニズムは、相補的標的ＲＮＡへのリボザイム分子の配列特異性ハイブリダイゼーションとその後に起こる内ヌクレオチド鎖切断に関

与している。例えば、遺伝子操作で作られたハンマーヘッド型リボザイム分子が、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。

【０１９５】

任意のＲＮＡ標的内の特異的リボザイム切断部位を、ＧＵＡ、ＧＵＵ、ＧＵＣ配列を含めたリボザイム切断部位に対する標的分子をスキャンすることによって先ず同定する。一度同定すると、切断部位を含む標的遺伝子の領域に対応する１５～２０リボヌクレオチドの短いＲＮＡ配列が、そのオリゴヌクレオチドを機能不全にするような２次構造の特徴をもっていないかを評価することが可能になる。候補標的の適合性の評価も、リボヌクレアーゼ保護アッセイを用いて相補的オリゴヌクレオチドとのハイブリダイゼーションのアクセス可能性をテストすることによって行うことができる。

【０１９６】

本発明の相補リボ核酸分子及びリボザイムは、核酸分子合成のために当分野でよく知られている任意の方法を用いて作製し得る。任意の方法には、固相ホスホラミダイト化学合成のようにオリゴヌクレオチドを化学的に合成する方法がある。或いは、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等の好適なＲＮＡポリメラーゼプロモーターを用いて多様なベクター内に取り込むことが可能である。或いは、相補的ＲＮＡを構成的或いは誘導的に合成するようなこれらｃＤＮＡ産物を、細胞系、細胞または組織内に導入することができる。

【０１９７】

ＲＮＡ分子を修飾して細胞内の安定性を高め、半減期を長くすることができる。限定するものではないが可能な修飾には、分子の５’末端、３’末端、あるいはその両方においてフランキング配列を追加したり、分子の主鎖内においてホスホジエステラーゼ結合ではなくホスホロチオネートまたは２’Ｏ－メチルを使用したりすることが含まれる。この概念は、ＰＮＡの産出に固有のものであり、これら全ての分子に拡大することができる。

それには、内在性エンドヌクレアーゼによって容易には認識されないアデニン、シチジン、グアニン、チミン、及びウリジンにアセチル－、メチル－、チオ－及び同様の修飾をしたものの他、非従来型塩基、例えばイノシン、クエオシン（ｑｕｅｏｓｉｎｅ）、ワイブトシン（ｗｙｂｕｔｏｓｉｎｅ）等を加えることでできる。

【０１９８】

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。特異ポリヌクレオチドの発現変異を起こさせるのに有効な化合物には、限定するものではないが、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。

【０１９９】

特異ポリヌクレオチドの変異発現における有効性に対して、少なくとも１個から複数個の試験化合物をスクリーニングし得る。試験化合物は、当分野で通常知られている任意の方法により得られる。 このような方法には、ポリヌクレオチドの発現を変異させる場合と、既存の市販または専売の、天然または非天然の化合物ライブラリから選択する場合と、標的ポリヌクレオチドの化学的及び／または構造的特性に基づく化合物を合理的にデザインする場合と、組合せ的にまたは無作為に生成した化合物のライブラリから選択する場合に有効であることが知られているような化合物の化学修飾がある。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少なくとも１つの試験化合物に曝露して得る。サンプルには例えば、無傷細胞、透過化処理した細胞、ｉｎ ｖｉｔｒｏ の無細胞系または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変異する際に試験化合物が有効であることを示している。特異ポリヌクレオチドの変異発現に有効な化合物に対して、例えば分裂酵母（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を調べるためのオリゴヌクレオチド（デオキシリボヌクレオチド、リボヌクレオチド、ペプチド核酸、修飾オリゴヌクレオチド）の組み合わせライブラリをスクリーニングすることに関与している（Ｂｒｕｉｃｅ，Ｔ．Ｗ．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）。

【０２００】

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

【０２０１】

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。

【０２０２】

本発明の追加実施例は、通常薬剤として許容できる賦形剤で処方される活性成分を含む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、ゴム及びタンパク質がある。様々な処方が通常知られており、詳細はＲｅｍｉｎｇｔｏｎ’ｓ Ｐｈａｒｍａｃｅｕｔｉｃａｌ Ｓｃｉｅｎｃｅｓ（Ｍａａｃｋ Ｐｕｂｌｉｓｈｉｎｇ，Ｅａｓｔｏｎ ＰＡ）の最新版に記載されている。このような組成物は、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。

【０２０３】

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。

【０２０４】

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胞領域を介しての肺送達が最近向上したことにより、インスリン等の薬剤を実質的に血液循環へ輸送することを

可能にした（Ｐａｔｔｏｎ，　Ｊ．Ｓ．　ら，　米国特許第５，９９７，８４８号等を参照）。肺送達は、針注射なしに投与する点で優れており、有毒な可能性のある浸透エンハンサーの必要性をなくす。

【０２０５】

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成分が含まれる。有効投与量の決定は、当業者の能力の範囲内で行う。

【０２０６】

ＩＧＳＦＰ　またはその断片を含む高分子を直接細胞内に送達するべく、特殊な形態に組成物が調製されるのが好ましい。例えば、細胞不透過性高分子を含むリポソーム製剤は、細胞融合及び高分子の細胞内送達を促進し得る。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Ｎ末端部に結合することもできる。このようにして生成された融合タンパク質は、マウスモデル系の脳を含む全ての組織の細胞に形質導入することがわかっている（Ｓｃｈｗａｒｚｅ，　Ｓ．Ｒ．　他　（１９９９）　Ｓｃｉｅｎｃｅ　２８５：１５６９－１５７２）。

【０２０７】

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

【０２０８】

医学的に効果的な薬用量は、症状や容態を回復させる、たとえば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処方成分の量に関連する。治療有効性及び毒性は、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ED_{50}（集団の５０％の治療有効量）またはLD_{50}（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。治療効果に対する毒性効果の用量比は、治療指数であり、LD_{50}/ED_{50}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を処方するのに用いられる。このような組成物が含まれる投与量は、毒性を殆ど或いは全く含まず、ED_{50}を含むような血中濃度の範囲にあることが好ましい。用いられる投与形態、患者の感受性及び投与の経路によって、投与量はこの範囲内で様々に変わる。

【０２０９】

正確な投与量は、治療が必要な被験者に関する要素を考慮して、現場の医者が決定することになる。効果的なレベルの活性成分を与え、或いは所望の効果を維持するべく、投与量及び投与を調節する。被験者に関する要素としては、疾患の重症度、患者の通常の健康状態、患者の年齢、体重及び性別、投与の時間及び頻度、薬剤の配合、反応感受性及び治療に対する応答等を考慮する。作用期間が長い組成物は、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。

【０２１０】

通常の投与量は、投与の経路にもよるが約０．１～１００，０００μｇであり、合計で約１ｇまでとする。特定の投与量及び送達方法に関するガイダンスは文献に記載されており、現場の医者は通常それを利用することができる。 当業者は、タンパク質またはインヒビターに対する処方とは異なる、ヌクレオチドに対する処方を利用することになる。同様に、ポリヌクレオチドまたはポリペプチドの送達は、特定の細胞、状態、位置等に特異的なものとなる。

【０２１１】

（診断）

別の実施例では、ＩＧＳＦＰ　に特異的に結合する抗体が、ＩＧＳＦＰ　の発現によって特

徴付けられる疾患の診断、またはＩＧＳＦＰ やＩＧＳＦＰ 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作製される。ＩＧＳＦＰ の診断アッセイには、抗体及び標識を用いてヒトの体液或いは細胞や組織から採取されたものからＩＧＳＦＰ 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。

【０２１２】

ＩＧＳＦＰを測定するための様々なプロトコル、例えばＥＬＩＳＡ、ＲＩＡ、ＦＡＣＳ等が本技術分野において知られており、ＩＧＳＦＰ発現の修正レベル或いは異常レベルを診断する基準を提供する。ＩＧＳＦＰ 発現の正常値或いは標準値は、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ＩＧＳＦＰ 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ＩＧＳＦＰ の発現の量が基準値と比較される。標準値と被験者との偏差が疾患を診断するパラメータとなる。

【０２１３】

本発明の別の実施例によれば、ＩＧＳＦＰ 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ＩＧＳＦＰ 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、ＩＧＳＦＰ の存在の有無、更に過剰な発現を調べ、治療中のＩＧＳＦＰ 値の調節を監視する。

【０２１４】

一実施形態では、ＩＧＳＦＰ 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、ＩＧＳＦＰ をコードする核酸配列を同定することが可能である。プローブが、例えば５’調節領域である高度に特異的な領域から作られていようと、あるいは、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ようと、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によって、プローブがＩＧＳＦＰをコードする天然の配列のみ、或いは対立遺伝子や関連配列を同定するかどうかが決まるであろう。

【０２１５】

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：７－１２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。

【０２１６】

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、^{32}Ｐまたは^{35}Ｓ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。

【０２１７】

ＩＧＳＦＰをコードするポリヌクレオチド配列を用いて、ＩＧＳＦＰの発現に関連する疾患を診断することが可能である。そのような疾患の例として、限定するものではないが、免疫疾患の中には、後天性免疫不全症候群（ＡＩＤＳ）、ＢｒｕｔｏｎのＸ染色体連関無ガンマグロブリン血症、分類不能型免疫不全（ＣＶＩ）、デジョージ症候群（胸腺発育不全）、胸腺形成異常、分離型ＩｇＡ欠損症、重症複合型免疫不全疾患（ＳＣＩＤ）、血小板減少症および湿疹を伴う免疫不全（ヴィスコット‐オールドリッチ症候群、Ｃｈｅ´ｄｉａｋ‐東症候群、慢性肉芽腫症、遺伝性血管〔運動〕神経性水腫（浮腫）、クッシング病合併免疫不全症、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、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫性多腺性内分泌カンジダ性外胚葉ジストロフィー（ＡＰＥＣＥＤ）、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、糖尿病、肺気腫、リンパ球毒素性一時性リンパ球減少症、赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性大腸症候群、多発性硬化症、重症筋無力症、心筋または心膜炎症、骨関節炎、骨粗しょう症、膵炎、多発性筋炎、乾癬、ライター症候群、リウマチ様関節炎、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、血小板減少性紫斑病、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌合併症、血液透析、体外循環、ウイルス感染症、細菌感染症、真菌感染症、寄生虫感染症、原虫感染症、蠕虫感染症、外傷が含まれ、神経系疾患の中には、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、パーキソン病及びその他の錐体外路障害、筋萎縮性側策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性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、及びトゥーレット病と、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が含まれ、また発達障害も含まれ、そ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、Ｓｙｎｄｅｎｈａｍ舞踏病（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、筋肉疾患には、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、筋緊張性ジストロフィー、中核疾患、ネマリンミオパシー、橋中心髄鞘崩壊症、脂質筋疾患、ミトコンドリアミオパシー、感染性筋炎、多発性筋炎、皮膚性筋炎、封入体筋炎、甲状腺中毒筋炎およびエタノールミオパシーが含まれ、また、細胞増殖異常では日光性角化症及びアテローム性動脈硬化、滑液包炎、硬変、肝炎、混合型結合組織病（ＭＣＴＤ）、骨髄線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症、並びに腺癌及び白血病、リンパ腫、黒色腫、骨髄腫、肉腫、及び奇形癌、具体的には、副腎、膀胱、骨

、骨髄、脳、乳房、頚部、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。変異ＩＧＳＦＰの発現を検出するために、患者から採取した体液或いは組織を利用して、ＩＧＳＦＰをコードするポリヌクレオチド配列を、サザーン法やノーザン法、ドットブロット法、或いはその他の膜系の技術、ＰＣＲ法、ディップスティック（ｄｉｐｓｔｉｃｋ）、ピン（ｐｉｎ）、ＥＬＩＳＡ式アッセイ、及びマイクロアレイに使用することが可能である。このような定性方法または定量方法は、当分野で公知である。

【０２１８】

ある実施態様では、ＩＧＳＦＰ 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。ＩＧＳＦＰ をコードするヌクレオチド配列は、標準的な方法で標識化され、ハイブリダイゼーション複合体の形成に好適な条件下で、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であろう。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。患者のサンプルのシグナルの量が、対照サンプルと較べて著しく変わっている場合は、サンプル内のＩＧＳＦＰ をコードするヌクレオチド配列の変異レベルにより、関連する疾患の存在が明らかになる。このようなアッセイは、動物実験、臨床試験における特定の治療効果を推定するため、或いは個々の患者の治療をモニターするために用いることもできる。

【０２１９】

ＩＧＳＦＰ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下で、動物或いはヒトの何れかの正常な被験者から抽出された体液或いは細胞抽出物と、ＩＧＳＦＰ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得る。実質的に精製されたポリヌクレオチドを既知量用いて行った実験から得た値を正常な対象から得た値と比較することにより、標準ハイブリダイゼーションを定量することができる。このようにして得た標準値は、疾患の徴候を示す患者から得たサンプルから得た値と比較することができる。標準値からの偏差を用いて疾患の存在を確定する。

【０２２０】

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

【０２２１】

癌に関しては、個体からの生検組織における異常な量の転写物（過少発現または過剰発現）の存在は、疾患の発生素質を示したり、実際に臨床的症状が現れる前に疾患を検出する方法を提供したりし得る。この種のより明確な診断により、医療の専門家が予防方法また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。

【０２２２】

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、或いは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、最適な条件下で、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。

【０２２３】

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（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。ｆＳＣＣＰ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シークエンシング機などの高処理機器でアンプリマー（ａｍｐｌｉｍｅｒ）の検出が可能になる。更に、インシリコ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に配列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の実験室での調整及び統計モデル及びＤＮＡ配列クロマトグラムの自動分析を用いたシークエンシングのエラーに起因する配列の変異をフィルタリングして除去する。別の態様では、例えば高処理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。

【０２２４】

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

【０２２５】

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

【０２２６】

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、薬剤－標的相互作用及び遺伝子発現プロフィールをモニターまたは測定することが可能である。

【０２２７】

或る実施例は、或る組織または細胞タイプの転写イメージを生成するような本発明のポリヌクレオチドの使用に関連する。転写イメージは、特定の組織または細胞タイプにより遺伝子発現の包括的パターンを表す。包括的遺伝子発現パターンは、所定の条件下で所定の時間に発現した遺伝子の数及び相対存在量を定量することにより分析される（Ｓｅｉｌｈａｍｅｒ 他、米国特許第５，８４０，４８４号の「Ｃｏｍｐａｒａｔｉｖｅ Ｇｅｎｅ

Ｔｒａｎｓｃｒｉｐｔ Ａｎａｌｙｓｉｓ」を参照。これらを引用することを以って本明細書の一部とする）。従って、特定の組織または細胞タイプの転写または逆転写全体に本発明のポリヌクレオチドまたはその補体をハイブリダイズすることにより、転写イメージを生成し得る。或る実施例では、本発明のポリヌクレオチドまたはその補体がマイクロアレイ上のエレメントのサブセットを複数含むような高処理フォーマットでハイブリダイゼーションを発生させる。結果として得られる転写イメージは、遺伝子活性のプロフィールを提供し得る。

【０２２８】

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

【０２２９】

本発明のポリヌクレオチドの発現のプロフィールを作製する転写イメージはまた、工業的または天然の環境化合物の毒性試験のみならず、ｉｎ ｖｉｔｒｏモデル系及び薬剤の前臨床評価に関連して使用し得る。全ての化合物は、作用及び毒性のメカニズムを暗示する、しばしば分子フィンガープリントまたは毒性シグネチャと称されるような特徴的な遺伝子発現パターンを惹起する（Ｎｕｗａｙｓｉｒ，Ｅ．Ｆ．他（１９９９）Ｍｏｌ．Ｃａｒｃｉｎｏｇ．２４：１５３－１５９、Ｓｔｅｉｎｅｒ，Ｓ．およびＮ．Ｌ．Ａｎｄｅｒｓｏｎ（２０００）Ｔｏｘｉｃｏｌ．Ｌｅｔｔ．１１２－１１３：４６７－４７１、該文献は特に引用することを以って本明細書の一部となす）。試験化合物が、既知の毒性を有する化合物のシグネチャと同一のシグネチャを有する場合には、毒性特性を共有している可能性がある。フィンガープリントまたはシグネチャは、多数の遺伝子及び遺伝子ファミリーからの発現情報を含んでいる場合は、最も有用且つ正確である。理想的には、発現をゲノム全域にわたって測定すると、最高品質のシグネチャが得られる。たとえ、発現が任意の試験された化合物によって変化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標準化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。規準化手順は、異なる化合物で処理した後の発現データの比較に有用である。毒性シグネチャの要素に遺伝子機能を割り当てることが毒性メカニズムの解釈に役立つが、毒性の予測につながるシグネチャを統計的に一致させるのに遺伝子機能の知識は必要とされない（例えば２０００年２月２９日に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た中毒学的スクリーニングの際に、全ての発現した遺伝子配列を含めることは重要且つ望ましいことである。

【０２３０】

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

【０２３１】

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

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

【０２３２】

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。或る実施例では、マイクロアレイ上でエレメントとして抗体を用い、マイクロアレイをサンプルに曝して各アレイ要素へのタンパク質結合レベルを検出することによりタンパク質発現レベルを定量する（Ｌｕｅｋｉｎｇ，Ａ．他（１９９９）Ａｎａｌ．Ｂｉｏｃｈｅｒｎ．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、サンプル中のタンパク質をチオールまたはアミノ反応性蛍光化合物と反応させ、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。

【０２３３】

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

【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。処理された生物学的サンプル中で発現したタンパク質は、各タンパク質の量を定量し得るように分離する。各タンパク質の量を、非処理生物学的サンプル中の対応するタンパク質の量と比較する。両サンプルのタンパク質量の差は、処理サンプル中の試験化合物に対する反応を示す。個々のタンパク質は、個々のタンパク質のアミノ酸残基をシークエンシングし、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。

【０２３５】

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

【０２３６】

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

分析する（Ｂｒｅｎｎａｎ，Ｔ．Ｍ．他（１９９５）の米国特許第５，４７４，７９６号、Ｓｃｈｅｎａ，Ｍ．他（１９９６）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ９３：１０６１４－１０６１９、Ｂａｌｄｅｓｃｈｗｅｉｌｅｒ他の（１９９５）ＰＣＴ出願第ＷＯ９５／２５１１１６号、Ｓｈａｌｏｎ，Ｄ．他の（１９９５）ＰＣＴ出願第ＷＯ９５／３５５０５号、Ｈｅｌｌｅｒ，Ｒ．Ａ．他（１９９７）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ９４：２１５０－２１５５、Ｈｅｌｌｅｒ，Ｍ．Ｊ．他の（１９９７）米国特許第５，６０５，６６２号等を参照）。様々なタイプのマイクロアレイが公知であり、詳細については、ＤＮＡ Ｍｉｃｒｏａｒｒａｙｓ：Ａ Ｐｒａｃｔｉｃａｌ Ａｐｐｒｏａｃｈ，Ｍ．Ｓｃｈｅｎａ，編集．（１９９９）Ｏｘｆｏｒｄ Ｕｎｉｖｅｒｓｉｔｙ Ｐｒｅｓｓ，Ｌｏｎｄｏｎに記載されている。該文献は、特に引用することを以って本明細書の一部となす。

【０２３７】

本発明の別の実施例ではまた、ＩＧＳＦＰをコードする核酸配列を用いて、天然のゲノム配列をマッピングするのに有用なハイブリダイゼーションプローブを作製することが可能である。コー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、ある場合には、コード配列より非コード配列が好ましい。例えば、多重遺伝子ファミリーのメンバー内でのコード配列の保存により、染色体マッピング中に望ましくないクロスハイブリダイゼーションが生じる可能性がある。核酸配列は、特定の染色体、染色体の特定領域または人工形成の染色体、例えば、ヒト人工染色体（ＨＡＣ）、酵母人工染色体（ＹＡＣ）、細菌人工染色体（ＢＡＣ）、細菌Ｐ１産物、或いは単一染色体ｃＤＮＡライブラリに対してマッピングされる。（例えば、Ｈａｒｒｉｎｇｔｏｎ，Ｊ．Ｊ．他（１９９７）Ｎａｔ．Ｇｅｎｅｔ．１５：３４５－３５５；Ｐｒｉｃｅ，Ｃ．Ｍ．（１９９３）Ｂｌｏｏｄ Ｒｅｖ．７：１２７１３４；ａｎｄ Ｔｒａｓｋ，Ｂ．Ｊ．（１９９１）Ｔｒｅｎｄｓ Ｇｅｎｅｔ．７：１４９１５４を参照。）一度マッピングすると、本発明の核酸配列を用いて例えば病状の遺伝を特定の染色体領域の遺伝または制限酵素断片長多型（ＲＦＬＰ）と相関させるような遺伝子連鎖地図を作製できる。（例えば、Ｌａｎｄｅｒ，Ｅ．Ｓ．ａｎｄ Ｄ．Ｂｏｔｓｔｅｉｎ（１９８６）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ８３：７３５３－７３５７を参照。）

【０２３８】

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相関し得る（前出のＨｅｉｎｚ－Ｕｌｒｉｃｈ，他（１９９５）ｉｎ Ｍｅｙｅｒｓ，９６５－９６８ページ等を参照）遺伝地図データの例は、種々の科学雑誌あるいはＯｎｌｉｎｅ Ｍｅｎｄｅｌｉａｎ Ｉｎｈｅｒｉｔａｎｃｅ ｉｎ Ｍａｎ（ＯＭＩＭ）のウェブサイトに見ることができる。物理的な染色体地図上のＩＧＳＦＰをコードする遺伝子の位置と特定の疾患との相関性、あるいは特定の疾患に対する素因との相関性は、この疾患と関連するＤＮＡの領域の決定に役立ち得るため、位置を決定するクローニングの作業を促進し得る。

【０２３９】

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

【０２４０】

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

【０２４１】

別の薬剤スクリーニング方法は、目的のタンパク質に対して好適な結合親和性を有する化合物を高い処理能力でスクリーニングするために用いられる（Ｇｅｙｓｅｎ，他（１９８４）ＰＣＴ ａｐｐｌｉｃａｔｉｏｎ ＷＯ８４／０３５６４等を参照。）この方法においては、多数の異なる小さな試験用化合物を固体基板上で合成する。試験用化合物は、ＩＧＳＦＰ或いはその断片と反応させ、洗浄する。次に、本技術分野でよく知られている方法で、結合した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に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。

【０２４２】

別の実施例では、ＩＧＳＦＰ と結合可能な中和抗体がＩＧＳＦＰ と結合するため試験用化合物と特に競合する、競合的薬剤スクリーニングアッセイを用いることができる。この方法では、抗体が、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。

【０２４３】

別の実施例では、将来に開発される分子生物学技術で、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）に依存しているならば、ＩＧＳＦＰをコードするヌクレオチド配列にその新技術を用い得る。

【０２４４】

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

【０２４５】

前述した及び以下に記載する全ての特許出願、特許、刊行物、特に米国特許出願第６０／２４９．６４５号に言及することをもって本明細書の一部とする。

【０２４６】

（実施例）

１　ｃＤＮＡライブラリの作製

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

【０２４７】

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。

【０２４８】

場合によってはＳｔｒａｔａｇｅｎｅ社へのＲＮＡ提供を行い、対応するｃＤＮＡライブラリをＳｔｒａｔａｇｅｎｅ社が作製することもあった。そうでない場合は、ＵＮＩＺＡＰベクターシステム（Ｓｔｒａｔａｇｅｎｅ）またはＳＵＰＥＲＳＣＲＩＰＴプラスミドシステム（Ｌｉｆｅ Ｔｅｃｈｎｏｌｏｇｉｅｓ）を用いて本技術分野で公知の推奨方法または類似の方法でｃＤＮＡを合成し、ｃＤＮＡライブラリを作製した（前出のＡｕｓｕｂｅｌ，１９９７，ｕｎｉｔ ５．１－６．６等を参照）。逆転写は、オリゴｄ（Ｔ）またはランダムプライマーを用いて開始した。合成オリゴヌクレオチドアダプターを二本鎖ｃＤＮＡに連結反応させ、好適な制限酵素でｃＤＮＡを消化した。殆どのライブラリに対して、ｃＤＮＡのサイズの選択（３００～１０００ｂｐ）は、ＳＥＰＨＡＣＲＹＬ Ｓ１０００、ＳＥＰＨＡＲＯＳＥ ＣＬ２ＢまたはＳＥＰＨＡＲＯＳＥ ＣＬ４Ｂカラムクロマトグラフィー（Ａｍｅｒｓｈａｍ Ｐｈａｒｍａｃｉａ Ｂｉｏｔｅｃｈ）、或いは調製用アガロースゲル電気泳動法を用いて行った。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に連結させた。 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 （Ｉｎｖｉｔｒｏｇｅｎ）、ＰＣＭＶ－ＩＣＩＳプラスミド（Ｓｔｒａｔａｇｅｎｅ）、ｐＩＧＥＮ （Ｉｎｃｙｔｅ Ｇｅｎｏｍｉｃｓ，Ｐａｌｏ Ａｌｔｏ ＣＡ）、またはｐＩＮＣＹ （Ｉｎｃｙｔｅ Ｇｅｎｏｍｉｃｓ）、あるいはその誘導体である。組換えプラスミドは、Ｓｔｒａｔａｇｅｎｅ社のＸＬ１－Ｂｌｕｅ、ＸＬ１－ＢＩｕｅＭＲＦまたはＳＯＬＲ、或いはＬｉｆｅ Ｔｅｃｈｎｏｌｏｇｉｅｓ社のＤＨ５α、ＤＨ１０ＢまたはＥＬＥＣＴＲＯＭＡＸ ＤＨ１０Ｂを含む大腸菌細胞に形質転換した。

【０２４９】

２　ｃＤＮＡクローンの単離

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

【０２５０】

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

【０２５１】

３　シークエンシング及び分析

実施例２に記載したようにプラスミドから回収したＩｎｃｙｔｅ ｃＤＮＡを、以下に示

すようにシークエンシングした。ｃＤＮＡのシークエンス反応は、標準的方法或いは高処理装置、例えばＡＢＩ ＣＡＴＡＬＹＳＴ ８００ サーマルサイクラー（Ａｐｐｌｉｅｄ Ｂｉｏｓｙｓｔｅｍｓ）またはＰＴＣ－２００ サーマルサイクラー（ＭＪ Ｒｅｓｅａｒｃｈ）をＨＹＤＲＡマイクロディスペンサー（Ｒｏｂｂｉｎｓ Ｓｃｉｅｎｔｉｆｉｃ）またはＭＩＣＲＯＬＡＢ ２２００（Ｈａｍｉｌｔｏｎ）液体転移システムと併用して処理した。ｃＤＮＡのシークエンス反応は、Ａｍｅｒｓｈａｍ Ｐｈａｒｍａｃｉａ Ｂｉｏｔｅｃｈ社が提供する試薬、またはＡＢＩシークエンシングキット、例えばＡＢＩ ＰＲＩＳＭ ＢＩＧＤＹＥ Ｔｅｒｍｉｎａｔｏｒ ｃｙｃｌｅ ｓｅｑｕｅｎｃｉｎｇ ｒｅａｄｙ ｒｅａｃｔｉｏｎ ｋｉｔ（Ａｐｐｌｉｅｄ Ｂｉｏｓｙｓｔｅｍｓ）の試薬を用いて準備した。ｃＤＮＡのシークエンス反応の電気泳動的分離及び標識したポリヌクレオチドの検出には、ＭＥＧＡＢＡＣＥ １０００ ＤＮＡシークエンシングシステム（Ｍｏｌｅｃｕｌａｒ Ｄｙｎａｍｉｃｓ）か、標準ＡＢＩプロトコル及び塩基対呼び出しソフトウェアを用いるＡＢＩ ＰＲＩＳＭ ３７３または３７７シークエンシングシステム（Ａｐｐｌｉｅｄ Ｂｉｏｓｙｓｔｅｍｓ）か、或いはその他の本技術分野でよく知られている配列解析システムを用いた。 ｃＤＮＡ配列内のリーディングフレームは、標準的方法（前出のＡｕｓｕｂｅｌ，１９９７，ｕｎｉｔ ７．７に概説）を用いて決定した。ｃＤＮＡ配列の幾つかを選択して、実施例８に記載した方法で配列を伸長させた。

【０２５２】

Ｉｎｃｙｔｅ ｃＤＮＡ配列に由来するポリヌクレオチド配列は、ベクター、リンカー及びポリ（Ａ）配列を除去し、あいまいな塩基対をマスクすることによって有効性を確認した。 その際、ＢＬＡＳＴ、動的プログラミング及び隣接ジヌクレオチド頻度分析に基づくアルゴリズム及びプログラムを用いた。次に、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）及びヒト、ラット、マウス、線虫、出芽酵母（Ｓａｃｃｈａｒｏｍｙｃｅ ｓｃｅｒｅｖｉｓｉａｅ）、分裂酵母（Ｓｃｈｉｚｏｓａｃｃｈａｒｏｍｙｃｅｓ ｐｏｍｂｅ）、鵞口瘡カンジダからの配列を含むＰＲＯＴＥＯＭＥデータベース（Ｉｎｃｙｔｅ Ｇｅｎｏｍｉｃｓ，Ｐａｌｏ Ａｌｔｏ ＣＡ）、並びにＰＦＡＭ等の隠れマルコフモデル（ＨＭＭ）ベースのタンパク質ファミリーデータベースに対して問い合わせた（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）Ｃｕｒｒ．Ｏｐｉｎ．Ｓｔｒｕｃｔ．Ｂｉｏｌ．６：３６１－３６５を参照のこと）。問合せは、ＢＬＡＳＴ、ＦＡＳＴＡ、ＢＬＩＭＰＳ及びＨＭＭＥＲに基づくプログラムを用いて行った。Ｉｎｃｙｔｅ ｃＤＮＡ配列は、完全長のポリヌクレオチド配列を産出するように構築した。或いは、ＧｅｎＢａｎｋ ｃＤＮＡ、ＧｅｎＢａｎｋ ＥＳＴ、ステッチされた配列、ストレッチされた配列またはＧｅｎｓｃａｎ予測コード配列（実施例４及び５を参照）を用いてＩｎｃｙｔｅ ｃＤＮＡの集団を完全長まで伸長させた。Ｐｈｒｅｄ、Ｐｈｒａｐ及びＣｏｎｓｅｄに基づくプログラムを用いて構築し、ＧｅｎＭａｒｋ、ＢＬＡＳＴ及びＦＡＳＴＡに基づくプログラムを用いてｃＤＮＡの集団をオープンリーディングフレームに対してスクリーニングした。対応する完全長ポリペプチド配列を誘導するべく完全長ポリヌクレオチド配列を翻訳した。或いは、本発明のポリペプチドは完全長翻訳ポリペプチドの任意のメチオニン残基で開始し得る。引き続いて、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Ｐｒｏｓｉｔｅ等のデータベース、並びに、ＰＦＡＭ等の隠れマルコフモデル（ＨＭＭ）ベースのタンパク質ファミリーデータベースに対する問合せによって完全長ポリペプチド配列を分析した。完全長ポリヌクレオチド配列はまた、ＭＡＣＤＮＡＳＩＳ ＰＲＯソフトウェア（日立ソフトウェアエンジニアリング，Ｓｏｕｔｈ Ｓａｎ Ｆｒａｎｃｉｓｃｏ ＣＡ）及びＬＡＳＥＲＧＥＮＥソフトウェア（ＤＮＡＳＴＡＲ）を用いて分析した。ポリヌクレオチド及びポリペプチド配列アラインメントは、アラインメントした配列と配列の一致率も計算するＭＥＧＡＬＩＧＮマルチシークエンスア

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作成する。

【０２５３】

Ｉｎｃｙｔｅ ｃＤＮＡ及び完全長配列の分析及びアセンブリに利用したツール、プログラム及びアルゴリズムの概略と、適用可能な説明、引用文献、閾値パラメータを表７に示す。用いたツール、プログラム及びアルゴリズムを表７の列１に、それらの簡単な説明を列２に示す。 列３は好適な引用文献であり、全ての文献はそっくりそのまま引用を以って本明細書の一部となす。 適用可能な場合には、列４は２つの配列が一致する強さを評価するために用いたスコア、確率値その他のパラメータを示す（スコアが高いほど、または確率値が低いほど、２配列間の同一性が高くなる）。

【０２５４】

完全長のポリヌクレオチド配列およびポリペプチド配列の組み立て及び分析に用いる上記のプログラムは、ＳＥＱ ＩＤ ＮＯ：７－１２のポリヌクレオチド配列断片の同定にも利用できる。 ハイブリダイゼーション及び増幅技術に有用である約２０～約４０００ヌクレオチドの断片を表４の列４に示した。

【０２５５】

４ ゲノムＤＮＡからのコード配列の同定及び編集

推定免疫グロブリンスーパファミリタンパク質は、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．およびＳ．Ｋａｒｌｉｎ（１９９７）Ｊ．Ｍｏｌ．Ｂｉｏｌ．２６８：７８－９４及びＢｕｒｇｅ，Ｃ．及びＳ．Ｋａｒｌｉｎ（１９９８）Ｃｕｆｆ．Ｏｐｉｎ．Ｓｔｒｕｃｔ．Ｂｉｏｌ．８：３４６－３５４参照）。プログラムは予測エキソンを連結し、メチオニンから停止コドンに及ぶ構築されたｃＤＮＡ配列を形成する。Ｇｅｎｓｃａｎの出力は、ポリヌクレオチド及びポリペプチド配列のＦＡＳＴＡデータベースである。Ｇｅｎｓｃａｎが一度に分析する配列の最大範囲は、３０ｋｂに設定した。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ファミリ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ファミリタンパク質について問合せて分析した。また、潜在的な免疫グロブリンスーパファミリタンパク質が、免疫グロブリンスーパファミリタンパク質として注釈が付けられたＩｎｃｙｔｅ ｃＤＮＡ配列に対する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のＧｅｎｓｃａｎにより予測された配列のエラーを修正する。ＢＬＡＳＴ分析はまた、任意のＩｎｃｙｔｅ ｃＤＮＡまたはＧｅｎｓｃａｎ予測配列の公共ｃＤＮＡ適用範囲の発見に用いられるので、転写の証拠を提供する。Ｉｎｃｙｔｅ ｃＤＮＡ適用範囲が利用できる場合には、この情報を用いてＧｅｎｓｃａｎ予測配列を修正または確認した。完全長ポリヌクレオチド配列は、実施例３に記載された構築プロセスを用いて、Ｉｎｃｙｔｅ ｃＤＮＡ配列及び／または公共のｃＤＮＡ配列でＧｅｎｓｃａｎ予測コード配列を構築することにより得た。或いは、完全長ポリヌクレオチド配列は編集または非編集のＧｅｎｓｃａｎ予測コード配列に完全に由来する。

【０２５６】

５ ｃＤＮＡ配列データを使ったゲノム配列データの構築

ステッチ配列（Ｓｔｉｃｈｅｄ Ｓｅｑｕｅｎｃｅ）

部分ｃＤＮＡ配列は、実施例４に記載のＧｅｎｓｃａｎ遺伝子同定プログラムにより予測されたエキソンを用いて伸長させた。実施例３に記載されたように構築された部分ｃＤＮＡは、ゲノムＤＮＡにマッピングし、関連するｃＤＮＡ及び１つ若しくは複数のゲノム配列から予測されたＧｅｎｓｃａｎエキソンを含むクラスタに分解した。ｃＤＮＡ及びゲノ

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

【０２５７】

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

【０２５８】

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：７－１２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：７－１２と一致するこれらのデータベースの配列を、Ｐｈｒａｐ（表７）などの構築アルゴリズムを使用して、連続及び重複した配列のクラスターに組み入れた。スタンフォード・ヒトゲノムセンター（ＳＨＧＣ）、ホワイトヘッド・ゲノム研究所（ＷＩＧＲ）、Ｇｅｎｅｔｈｏｎ等の公的な情報源から入手可能な放射線ハイブリッド及び遺伝地図データを用いて、クラスタ化された配列が前にマッピングされていたかを確認した。マッピングされた配列がクラスタに含まれている場合は、個々の配列番号を含めてそのクラスタの全配列が地図上の位置に割り当てられた。

【０２５９】

地図上の位置は、ヒト染色体の範囲または間隔として表される。センチモルガン間隔の地図上の位置は、染色体のｐアームの末端に関して測定する。（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、１ｃＭ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よって広範囲に変化する。）ｃＭ距離は、配列が各クラスタ内に含まれるような放射線ハイブリッドマーカーに対して境界を提供するようなＧｅｎｅｔｈｏｎによってマッピングされた遺伝マーカーに基づく。ＮＣＢＩ「ＧｅｎｅＭａｐ９

９」（ｈｔｔｐ：／／ｗｗｗ．ｎｃｂｉ．ｎｌｍ．ｎｉｈ．ｇｐｖ／ｇｅｎｅｍａｐ）などの一般個人が入手可能なヒト遺伝子マップおよびその他の情報源を用いて、上記した区間が既に同定されている疾患遺伝子マップ内若しくは近傍に位置するかを決定できる。

【０２６０】

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関与している。（前出のＳａｍｂｒｏｏｋ，７章、同Ａｕｓｕｂｅｌ．Ｆ．Ｍ．他，４章及び１６章等を参照。）

【０２６１】

ＢＬＡＳＴに適用する類似のコンピュータ技術を用いて、ＧｅｎＢａｎｋやＬｉｆｅＳｅｑ（Ｉｎｃｙｔｅ　Ｐｈａｒｍａｃｅｕｔｉｃａｌｓ）等のヌクレオチドデータベースにおいて同一または関連分子を検索する。ノーザン分析は、多数膜系ハイブリダイゼーションよりも非常に速い。更に、特定の一致を厳密な一致、或いは類似的な一致として分類するか否かを決定するため、コンピュータ検索の感度を変更することができる。検索の基準は積スコアであり、次式で定義される。

【０２６２】

【数１】

$$\frac{(\text{BLAST スコア} \times \text{配列一致率})}{5 \times (\text{長さ(配列１)，長さ(配列２)})\text{の最小値}}$$

積スコアは、２つの配列間の類似度及び配列が一致する長さの両方を考慮している。積スコアは、０～１００の規準化された値であり、次のようにして求める。ＢＬＡＳＴスコアにヌクレオチドの配列一致率を乗じ、その積を２つの配列の短い方の長さの５倍で除する。高スコアリングセグメント対（ＨＳＰ）に一致する各塩基に＋５のスコアを割り当て、各不適性塩基対に－４を割り当てることにより、ＢＬＡＳＴスコアを計算する。２つの配列は、２以上のＨＳＰを共有し得る（ギャップにより隔離され得る）。２以上のＨＳＰがある場合には、最高ＢＬＡＳＴスコアの塩基対を用いて積スコアを計算する。積スコアは、断片的オーバーラップと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。

【０２６３】

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

【０２６４】

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もまた、完全長分子の適切な断片から設計したオリゴヌクレオチドプライマーを用いて該断片を伸長させて生成した。一方のプライマーは既知の断片の５’伸長を開始するべく合成し、他方のプライマーは既知の断片の３’伸長を開始するべく合成した。開始プライマーは、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、ＯＬＩＧＯ４．０６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。

【０２６５】

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。

【０２６６】

高忠実度の増幅が、当業者によく知られている方法を利用したＰＣＲ法によって得られた。ＰＣＲは、ＰＴＣ－２００サーマルサイクラー（ＭＪ　Ｒｅｓｅａｒｃｈ，Ｉｎｃ．）を用いて９６穴プレート内で行った。反応混合液は、鋳型ＤＮＡ及び２００ｎｍｏｌの各プライマー、Mg^{2+}と$(NH_4)_2SO_4$と２－メルカプトエタノールを含むバッファー、Ｔａｑ　ＤＮＡポリメラーゼ（Ａｍｅｒｓｈａｍ　Ｐｈａｒｍａｃｉａ　Ｂｉｏｔｅｃｈ）、ＥＬＯＮＧＡＳＥ酵素（Ｌｉｆｅ　Ｔｅｃｈｎｏｌｏｇｉｅｓ）、Ｐｆｕ　ＤＮＡポリメラーゼ（Ｓｔｒａｔａｇｅｎｅ）を含む。プライマーの組、ＰＣＩ　ＡとＰＣＩ　Ｂに対して以下のパラメータで増幅を行った。ステップ１：９４℃，３分、ステップ２：９４℃，１５秒、ステップ３：６０℃，１分、ステップ４：６８℃，２分、ステップ５：ステップ２、３および４を２０回反復する。ステップ６：６８℃，５分、ステップ７：４℃で保存する。別法では、プライマーの組、Ｔ７とＳＫ＋に対して以下のパラメータで増幅を行った。ステップ１：９４℃，３分、ステップ２：９４℃，１５秒、ステップ３：５７℃，１分、ステップ４：６８℃，２分、ステップ５：ステップ２、３および４を２０回反復する。ステップ６：６８℃、５分、ステップ７：４℃で保存する。

【０２６７】

各ウェルのＤＮＡ濃度は、１Ｘ　ＴＥ及び０．５μｌの希釈していないＰＣＲ産物に溶解した１００μｌのＰＩＣＯＧＲＥＥＮ定量試薬（０．２５％（ｖ／ｖ）　ＰＩＣＯＧＲＥＥＮ；Ｍｏｌｅｃｕｌａｒ　Ｐｒｏｂｅｓ，Ｅｕｇｅｎｅ　ＯＲ）を不透明な蛍光光度計プレート（Ｃｏｍｉｎｇ　Ｃｏｓｔａｒ，Ａｃｔｏｎ　ＭＡ）の各ウェルに分配してＤＮＡが試薬と結合できるようにして測定する。サンプルの蛍光を計測してＤＮＡの濃度を定量するためにプレートをＦｌｕｏｒｏｓｋａｎ　ＩＩ（Ｌａｂｓｙｓｔｅｍｓ　Ｏｙ，Ｈｅｌｓｉｎｋｉ，Ｆｉｎｌａｎｄ）でスキャンした。反応混合物のアリコート５～１０μｌを１％アガロースミニゲル上で電気泳動法によって解析し、どの反応が配列の伸長に成功したかを決定した。

【０２６８】

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・シークエンシングのために、消化したヌクレオチドを低濃度（０．６～０．８％）のアガロースゲル上で分離し、断片を切除し、寒天をＡｇａｒ　ＡＣＥ（Ｐｒｏｍｅｇａ）で消化した。Ｔ４リガーゼ（Ｎｅｗ　Ｅｎｇｌａｎｄ　Ｂｉｏｌａｂｓ，Ｂｅｖｅｒｌｙ　ＭＡ）を用いて伸長させたクローンを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部分を満たし、大腸菌細胞に形質移入した。形質移入した細胞を選択して抗生物質を含む培地に移し、それぞれのコロニーを切りとってＬＢ／２Ｘカルベニシリン培養液の３８４ウェルプレートに３７℃で一晩培養した。

【０２６９】

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

【０２７０】

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

【０２７１】

９ 個々のハイブリダイゼーションプローブの標識及び使用

ＳＥＱ ＩＤ ＮＯ：７－１２由来のハイブリダイゼーションプローブを用いて、ｃＤＮＡ、ｍＲＮＡ、またはゲノムＤＮＡをスクリーニングする。約２０塩基対からなるオリゴヌクレオチドの標識について特に記載するが、より大きなヌクレオチド断片に対しても事実上同一の手順が用いられる。オリゴヌクレオチドは、ＯＬＩＧＯ ４．０６ソフトウェア（Ｎａｔｉｏｎａｌ Ｂｉｏｓｃｉｅｎｃｅｓ）等の最新ソフトウェアを用いて設計し、各オリゴマー５０ｐｍｏｌと、［γ－^{32}Ｐ］アデノシン３リン酸 （Ａｍｅｒｓｈａｍ Ｐｈａｒｍａｃｉａ Ｂｉｏｔｅｃｈ）２５０μＣｉと、Ｔ４ポリヌクレオチドキナーゼ（ＤｕＰｏｎｔ ＮＥＮ， Ｂｏｓｔｏｎ ＭＡ）を混合することにより標識する。標識したオリゴヌクレオチドは、ＳＥＰＨＡＤＥＸ Ｇ－２５超細繊分子サイズ排除デキストラン ビードカラム（Ａｍｅｒｓｈａｍ Ｐｈａｒｍａｃｉａ Ｂｉｏｔｅｃｈ）を用いて十分に精製する。Ａｓｅ Ｉ、Ｂｇｌ ＩＩ、Ｅｃｏ ＲＩ、Ｐｓｔ Ｉ、Ｘｂａ１またはＰｖｕ ＩＩ（ＤｕＰｏｎｔ ＮＥＮ）のいずれか１つのエンドヌクレアーゼで消化したヒトゲノムＤＮＡの典型的な膜ベースのハイブリダイゼーション解析において、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。

【０２７２】

各消化物から得たＤＮＡは、０．７％アガロースゲル上で分画してナイロン膜（Ｎｙｔｒａｎ Ｐｌｕｓ， Ｓｃｈｌｅｉｃｈｅｒ ＆ Ｓｃｈｕｅｌｌ， Ｄｕｒｈａｍ ＮＨ）に移す。ハイブリダイゼーションは、４０℃で１６時間行う。 非特異的シグナルを除去するため、例えば０．１×クエン酸ナトリウム食塩水及び０．５％ドデシル硫酸ナトリウムに一致する条件下で、ブロットを室温で順次洗浄する。オートラジオグラフィー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。

【０２７３】

１０ マイクロアレイ

マイクロアレイ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、圧電印刷（インクジェット印刷、前出のＢａｌｄｅｓｃｈｗｅｉｌｅｒ等を参照）、機械的マイクロスポッティング技術及びこれらから派生したものを用いて達成することが可能である。上記各技術において基板は、均一且つ非多孔性の固体とするべきである（Ｓｃｈｅｎａ（１９９９）前出）。推奨する基板には、シリコン、シリカ、スライドガラス、ガラスチップ及びシリコンウエハがある。或いは、ドットブロット法またはスロットブロット法に類似のアレイを利用して、熱的、紫外線的、化学的または機械的結合手順を用いて基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、当業者に周知の利用可能な方法や機械を用いて作製でき、任意の適正数のエレメントを有し得る（Ｓｃｈｅｎａ，Ｍ．他（１９９５）Ｓｃｉｅｎｃｅ２７０：４６７－４７０、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）。

【０２７４】

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

【０２７５】

組織または細胞サンプルの調製

グアニジウムチオシアネート法を用いて組織サンプルから全ＲＮＡを単離し、オリゴ（ｄＴ）セルロース法を用いてポリ（Ａ）$^{+}$ＲＮＡを精製する。各ポリ（Ａ）$^{+}$ＲＮＡサンプルは、ＭＭＬＶ逆転写酵素、０．０５ｐｇ／μｌのオリゴ（ｄＴ）プライマー（２１ｍｅｒ）、１Ｘ 第１鎖バッファ、０．０３ｕｎｉｔ／μｌのＲＮアーゼ阻害剤、５００μＭのｄＡＴＰ、５００μＭのｄＧＴＰ、５００μＭのｄＴＴＰ、４０μＭのｄＣＴＰ、４０μＭのｄＣＴＰ－Ｃｙ３（ＢＤＳ）またはｄＣＴＰ－Ｃｙ５（Ａｍｅｒｓｈａｍ Ｐｈａｒｍａｃｉａ Ｂｉｏｔｅｃｈ）を用いて逆転写する。逆転写反応は、ＧＥＭＢＲＩＧＨＴキット（Ｉｎｃｙｔｅ）を用いてポリ（Ａ）$^{+}$ＲＮＡ含有の２５体積ｍｌ内で行う。特異的対照ポリ（Ａ）$^{+}$ＲＮＡは、非コード酵母ゲノムＤＮＡからｉｎ ｖｉｔｒｏ転写により合成する。３７℃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、２．５ｍｌの０．５Ｍ水酸化ナトリウムで処理し、８５℃で２０分間インキュベートし、反応を停止させてＲＮＡを分解させる。サンプルは、２つの連続するＣＨＲＯＭＡ ＳＰＩＮ ３０ゲル濾過スピンカラム（ＣＬＯＮＴＥＣＨ Ｌａｂｏｒａｔｏｒｉｅｓ，Ｉｎｃ．（ＣＬＯＮＴＥＣＨ），Ｐａｌｏ Ａｌｔｏ ＣＡ）を用いて精製する。 混合した後、２つの反応サンプルは、１ｍｌのグリコーゲン（１ｍｇ／ｍｌ）、６０ｍｌの酢酸ナトリウム及び３００ｍｌの１００％エタノールを用いてエタノール沈殿させる。サンプルは次に、ＳｐｅｅｄＶＡＣ（Ｓａｖａｎｔ Ｉｎｓｔｒｕｍｅｎｔｓ Ｉｎｃ．，Ｈｏｌｂｒｏｏｋ ＮＹ）を用いて乾燥して仕上げ、１４μｌの５×ＳＳＣ／０．２％ＳＤＳ中で再懸濁する。

【０２７６】

マイクロアレイの調製

本発明の配列を用いて、アレイエレメントを生成する。各アレイエレメントは、クローン

化ｃＤＮＡインサートによりベクター含有細菌性細胞から増幅する。ＰＣＲ増幅は、ｃＤＮＡインサートの側面に位置するベクター配列に相補的なプライマーを用いる。３０サイクルのＰＣＲで１～２ｎｇの初期量から５μｇ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。

【０２７７】

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グラス上に固定する。顕微鏡スライドグラス（Ｃｏｒｎｉｎｇ）は、処理中及び処理後に０．１％のＳＤＳ及びアセトン中で超音波をかけ、蒸留水で非常に良く洗って洗浄する。スライドグラスは、４％フッ化水素酸（ＶＷＲ Ｓｃｉｅｎｔｉｆｉｃ Ｐｒｏｄｕｃｔｓ Ｃｏｒｐｏｒａｔｉｏｎ（ＶＷＲ），Ｗｅｓｔ Ｃｈｅｓｔｅｒ ＰＡ）中でエッチングし、蒸留水中で広範囲にわたって洗浄し、９５％エタノール中で０．０５％アミノプロピルシラン（Ｓｉｇｍａ）を用いてコーティングする。コーティングしたスライドガラスは、１１０℃のオブンで硬化させる。

【０２７８】

米国特許第５，８０７，５２２号で説明されている方法を用いて、コーティングしたガラス基板にアレイエレメントを付加する。該特許は、引用を以って本明細書の一部となす。平均濃度が１００ｎｇ／μｌのアレイエレメントＤＮＡ１μｌを高速機械装置により開放型キャピラリープリンティングエレメント（ｏｐｅｎ ｃａｐｉｌｌａｒｙ ｐｒｉｎｔｉｎｇ ｅｌｅｍｅｎｔ）に充填する。装置はここで、スライド毎に約５ｎｌのアレイエレメントサンプルを加える。

【０２７９】

マイクロアレイには、ＳＴＲＡＴＡＬＩＮＫＥＲ ＵＶ架橋剤（Ｓｔｒａｔａｇｅｎｅ）を用いてＵＶ架橋する。マイクロアレイは、室温において０．２％ＳＤＳで１度洗浄し、蒸留水で３度洗浄する。リン酸緩衝生理食塩水（ＰＢＳ）（Ｔｒｏｐｉｘ，Ｉｎｃ．，Ｂｅｄｆｏｒｄ ＭＡ）中の０．２％カゼイン中において６０℃で３０分間マイクロアレイをインキュベートした後、前に行ったように０．２％ＳＤＳ及び蒸留水で洗浄することにより、非特異結合部位をブロックする。

【０２８０】

ハイブリダイゼーション

ハイブリダイゼーション反応液は、５×ＳＳＣ、０．２％ＳＤＳハイブリダイゼーション緩衝液にＣｙ３及びＣｙ５標識したｃＤＮＡ合成産物を各０．２μｇ含む９μｌのサンプル混合体を含めたものである。サンプル混合体は、６５℃まで５分間加熱し、マイクロアレイ表面上で等分して１．８ｃｍ^2 のカバーガラスで覆う。アレイは、顕微鏡スライドより僅かに大きい空洞を有する防水チェンバーに移す。チェンバーのコーナーに１４０μｌの５×ＳＳＣを加えることにより、チェンバー内部を湿度１００％に保持する。アレイを含むチェンバーは、６０℃で約６．５時間インキュベートする。アレイは、第１洗浄緩衝液中（１×ＳＳＣ、０．１％ＳＤＳ）において４５℃で１０分間洗浄し、第２洗浄緩衝液中（０．１×ＳＳＣ）において４５℃で１０分間各々３度洗浄して乾燥させる。

【０２８１】

検出

レポーター標識されたハイブリダイゼーション複合体は、Ｃｙ３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，Ｓａｎｔａ Ｃｌａｒａ ＣＡ）を備えた顕微鏡で検出する。２０×顕微鏡対物レンズ（Ｎｉｋｏｎ，Ｉｎｃ．，Ｍｅｌｖｉｌｌｅ ＮＹ）を用いて、アレイ上に励起レーザ光の焦点を当てる。アレイを含むスライドを顕微鏡のコンピュータ制御のＸ－Ｙステージに置き、対物レンズを通過してラスタースキャンする。本実施例で用いた１．８ｃｍ×１．８ｃｍのアレイは、２０μｍの解像度でスキャンした。

【０２８２】

２つの異なるスキャンのうち、混合ガスマルチラインレーザは２つの蛍光色素を連続的に励起する。発光された光は、波長に基づき分離され、２つの蛍光色素に対応する２つの光電子増倍管検出器（ＰＭＴ Ｒ１４７７，Ｈａｍａｍａｔｓｕ Ｐｈｏｔｏｎｉｃｓ Ｓｙｓｔｅｍｓ，Ｂｒｉｄｇｅｗａｔｅｒ ＮＪ）に送られる。アレイと光電子増倍管間に設置された好適なフィルターを用いて、シグナルをフィルターする。用いる蛍光色素の最大発光の波長は、Ｃｙ３では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。

【０２８３】

スキャンの感度は通常、既知濃度のサンプル混合体に添加されるｃＤＮＡ対照種により生成されるシグナル強度を用いて較正する。アレイ上の特定の位置には相補的ＤＮＡ配列が含まれ、その位置におけるシグナルの強度を重量比１：１００，０００でハイブリダイゼーション種と相関させる。異なる源泉（例えば試験される細胞及び対照細胞など）からの２つのサンプルを、各々異なる蛍光色素で標識し、他と異なって発現した遺伝子を同定するために単一のアレイにハイブリダイズする場合には、２つの蛍光色素で較正するｃＤＮＡのサンプルを標識し、ハイブリダイゼーション混合液に各々等量を加えることによって較正を行う。

【０２８４】

光電子増倍管の出力は、ＩＢＭ互換性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－ディジタル（Ａ／Ｄ）変換ボード（Ａｎａｌｏｇ Ｄｅｖｉｃｅｓ，Ｉｎｃ．，Ｎｏｒｗｏｏｄ ＭＡ）を用いてディジタル化される。ディ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直線的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。

【０２８５】

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

【０２８６】

１１　相補的ポリヌクレオチド

ＩＧＳＦＰをコードする配列或いはその任意の一部に対して相補的な配列は、天然のＩＧＳＦＰの発現を検出または低下あるいは阻害するために用いられる。約１５～３０塩基対を含むオリゴヌクレオチドの使用について記すが、これより小さな或いは大きな配列の断片の場合でも本質的に同じ方法を用いることができる。Ｏｌｉｇｏ４．０６ソフトウェア（Ｎａｔｉｏｎａｌ Ｂｉｏｓｃｉｅｎｃｅｓ）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。

【０２８７】

１２　ＩＧＳＦＰの発現

ＩＧＳＦＰ の発現及び精製は、細菌若しくはウイルスを基にした発現系を用いて行うことができる。細菌でＩＧＳＦＰ が発現するために、抗生物質耐性及びｃＤＮＡの転写レベルを高める誘導性のプロモーターを含む好適なベクターにｃＤＮＡをサブクローニングする。このようなプロモーターには、ｌａｃオペレーター調節エレメントに関連するＴ５またはＴ７バクテリオファージプロモーター及びｔｒｐ－ｌａｃ（ｔａｃ）ハイブリッドプロモーターが含まれるが、これらに限定するものではない。組換えベクターを、ＢＬ２

１（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β－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ス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ポリヘドリン遺伝子を、相同組換え或いは転移プラスミドの媒介を伴う細菌の媒介による遺伝子転移のどちらかによって、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、強い多角体プロモーターによって高いレベルのｃＤＮＡの転写が行われる。組換えバキュロウイルスは、多くの場合はＳｐｏｄｏｐｔｅｒａ ｆｒｕｇｉｐｅｒｄａ（Ｓｆ９）昆虫細胞に感染に用いられるが、ヒト肝細胞の感染にも用いられることもある。後者の感染の場合は、バキュロウイルスの更なる遺伝的変更が必要になる。（Ｅｎｇｅｌｈａｒｄ．Ｅ．Ｋ．他（１９９４）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ ９１：３２２４－３２２７、Ｓａｎｄｉｇ，Ｖ．他（１９９６）Ｈｕｍ．Ｇｅｎｅ Ｔｈｅｒ．７：１９３７－１９４５．等を参照）。

【０２８８】

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）、またはＦＬＡＧや６－Ｈｉｓなどのペプチドエピトープ標識で合成された融合タンパク質となるため、未精製の細胞溶解物からの組換え融合タンパク質の親和性ベースの精製が素早く１回で行うことができる。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 からタンパク分解的に切断できる。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）を用いて免疫親和性精製を可能にする。６ヒスチジン残基が連続して伸長した６－Ｈｉｓは、金属キレート樹脂（ＱＩＡＧＥＮ）上での精製を可能にする。タンパク質の発現及び精製の方法は、前出のＡｕｓｕｂｅｌ（１９９５）１０章、１６章に記載されている。これらの方法で精製したＩＧＳＦＰ を直接用いて以下の実施例１６、及び１７のアッセイを行うことができる。

【０２８９】

１３ 機能的アッセイ

ＩＧＳＦＰ の機能は、哺乳動物細胞培養系において生理学的に高められたレベルでのＩＧＳＦＰ をコードする配列の発現によって評価する。 ｃＤＮＡを、ｃＤＮＡを高いレベルで発現する強いプロモーターを含む哺乳動物発現ベクターにサブクローニングする。選択ベクターには、ｐＣＭＶ ＳＰＯＲＴプラスミド（Ｌｉｆｅ Ｔｅｃｈｎｏｌｏｇｉｅｓ）及びｐＣＲ ３．１プラスミド（Ｉｎｖｉｔｒｏｇｅｎ）が含まれ、どちらもサイトメガロウイルスプロモーターを有する。リポソーム製剤或いは電気穿孔法を用いて、５～１０μｇの組換えベクターをヒト細胞株、例えば内皮由来または造血由来の細胞株に一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１～２μｇのプラスミドを同時に形質移入する。標識タンパク質の発現により、形質移入細胞と非形質移入細胞を区別する手段が与えられる。 また、標識タンパク質の発現によって、ｃＤＮＡの組換えベクターからの発現を正確に予想できる。標識タンパク質は、例えば緑色蛍光タンパク質（ＧＦＰ；Ｃｌｏｎｔｅｃｈ）、ＣＤ６４またはＣＤ６４－ＧＦＰ融合タンパク質から選択できる。自動化された、レーザー光学に基づく技術であるフローサイトメトリー（ＦＣＭ）を用いて、ＧＦＰまたはＣＤ６４－ＧＦＰを発現する形質移入された細胞を同定し、その細胞のアポトーシス状態や他の細胞特性を評価する。ＦＣＭは、細胞死に先行するか或いは同時に発生する現象を診断する蛍光分子の取込を検出して計量する。このような現象として挙げられるのは、プロピジウムヨウ化物によるＤＮＡ染色によって計測される核ＤＮＡ内容物の変化、前方光散乱と９０°側方光散乱によって計測される細胞サイズと顆粒状性の変化、ブ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ン

ク質の発現の変化、及び蛍光複合アネキシンＶタンパク質の細胞表面への結合によって計測される原形質膜組成の変化とがある。フローサイトメトリー法については、Ｏｒｍｅｒｏｄ，Ｍ．Ｇ．（１９９４）Ｆｌｏｗ Ｃｙｔｏｍｅｔｒｙ Ｏｘｆｏｒｄ，Ｎｅｗ Ｙｏｒｋ，ＮＹ．に記述がある。

【０２９０】

遺伝子発現における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、形質転換された細胞表面で発現し、ヒト免疫グロブリンＧ（ＩｇＧ）の保存領域と結合する。形質転換された細胞と形質転換されない細胞とは、ヒトＩｇＧか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。

【０２９１】

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

【０２９２】

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。

【０２９３】

１５　特異的抗体を用いる天然ＩＧＳＦＰ の精製

天然または組換えＩＧＳＦＰ を、ＩＧＳＦＰ 特異抗体を用いたイムノアフィニティークロマトグラフィにより実質的に精製する。イムノアフィニティーカラムは、ＣＮＢｒ－活性化ＳＥＰＨＡＲＯＳＥ（Ａｍｅｒｓｈａｍ Ｐｈａｒｍａｃｉａ Ｂｉｏｔｅｃｈ）のような活性化クロマトグラフィー用レジンと抗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樹脂をブロックし、洗浄する。

【０２９４】

ＩＧＳＦＰ を含む培養液をイムノアフィニティーカラムに通し、ＩＧＳＦＰ を優先的に吸着できる条件で（例えば、界面活性剤の存在下において高イオン強度のバッファーで）そのカラムを洗浄する。抗体とＩＧＳＦＰ の結合を破壊する条件（例えばｐＨ２～３の緩衝液、或いは尿素またはチオシアン酸塩イオン等の高濃度のカオトロープ剤）でカラムを溶出させ、ＩＧＳＦＰ を回収する。

【０２９５】

１６　ＩＧＳＦＰと相互作用する分子の同定

ＩＧＳＦＰまたはその生物学的に活性である断片を、^{125}Ｉ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。

【０２９６】

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。

【０２９７】

ＩＧＳＦＰ はまた、ハイスループットな方法で酵母２ハイブリッドシステムを使用するＰＡＴＨＣＡＬＬＩＮＧプロセス（ＣｕｒａＧｅｎ Ｃｏｒｐ．，Ｎｅｗ Ｈａｖｅｎ ＣＴ）に用いて、遺伝子の２大ライブラリにコードされる遺伝子間の全ての相互作用を決定することができる（Ｎａｎｄａｂａｌａｎ，Ｋ．他（２０００）米国特許第６，０５７，１０１号）。（２０００）米国特許第６，０５７，１０１号）。

【０２９８】

１７ ＩＧＳＦＰ の活性の実証

ＩＧＳＦＰ 活性用アッセイによってＩＧＳＦＰ の血清からの抗原を認識し、沈澱させる能力が測定される。この活性は定量沈降反応によって測定される（Ｇｏｌｕｂ，Ｅ．Ｓ．他（１９８７）Ｉｍｍｕｎｏｌｏｇｙ：Ａ Ｓｙｎｔｈｅｓｉｓ，Ｓｉｎａｕｅｒ Ａｓｓｏｃｉａｔｅｓ，Ｓｕｎｄｅｒｌａｎｄ ＭＡ，１１３－１１５ページ）．ＩＧＳＦＰ は当分野において既知の方法を用いて同位体で標識される。種々の血清濃度で一定量の標識ＩＧＳＦＰ に加える。ＩＧＳＦＰ抗原複合物が溶液に沈殿し、これを遠心分離によって回収する。沈殿性ＩＧＳＦＰ抗原複合物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ＩＧＳＦＰ抗原複合物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ＩＧＳＦＰ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、沈澱性ＩＧＳＦＰ抗原複合物の量は、抗原の制限量と過剰量に対する感受性を特徴とするＩＧＳＦＰ活性の測定方法となる。

【０２９９】

別な方法として、ＩＧＳＦＰ 活性に対するアッセイは、細胞表面におけるＩＧＳＦＰ の発現を測定する。ＩＧＳＦＰ をコードするｃＤＮＡを、非白血球細胞系に形質移入する。細胞表面タンパク質をビオチンで標識する（ｄｅ ｌａ Ｆｕｅｎｔｅ，Ｍ．Ａ．他（１９９７）Ｂｌｏｏｄ ９０：２３９８－２４０５）。ＩＧＳＦＰ特異抗体を用いて免疫沈降を実行し、ＳＤＳ－ＰＡＧＥおよび免疫ブロット技術を用いて免疫沈降サンプルを分析する。標識された免疫沈降と未標識免疫沈降の比は、細胞表面に発現したＩＧＳＦＰ の量に比例する。

【０３００】

代替例として、ＩＧＳＦＰ 活性のアッセイは、ＩＧＳＦＰ 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ＩＧＳＦＰ 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）等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癒着（塊化）量を比較する。細胞の癒着量によりＩＧＳＦＰ の活性を直接的に測定できる。

【０３０１】

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

【０３０２】

（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。

【０３０３】

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

【０３０４】

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

【０３０５】

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。

【０３０６】

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。

【０３０７】

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。

【０３０８】

表７は、本発明のポリヌクレオチドおよび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。

【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
6471113	1	6471113CD1	7	6471113CB1
7701502	2	7701502CD1	8	7701502CB1
5727379	3	5727379CD1	9	5727379CB1
5831801	4	5831801CD1	10	5831801CB1
7195090	5	7195090CD1	11	7195090CB1
7473826	6	7473826CD1	12	7473826CB1

【表２】

表 2

ポリペプチド SEQ ID NO:	Incyte ポリペプチド ID	GenBank ID NO:	確率 スコア	GenBank 相同体
1	6471113CD1	g3068592	0.0	Purc [マウス] (Salbaum, J.M. (1998) Mech. Dev. 71:201-204)
2	7701502CD1	g6746556	8.8e-74	シアル酸結合免疫グロブリン様レクチン-8 [ヒト] Floyd, H. 他 (2000) J. Biol. Chem. 275:861-866)
3	5727379CD1	g9837568	7.1e-43	神経細胞ロイシンリッチリピートタンパク質-3[ラット]
4	5831801CD1	g312590	3.4e-15	胆汁糖タンパク質[マウス] (McCuaig, K. 他 (1993) Gene 127:173-183)
5	7195090CD1	g7290075	0.0	sdk遺伝子産物[キイロショウジョウバエ] (Adams, M.D. 他 (2000) Science 287:2185-2195)

【表３】

表3－1

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基	潜在的リン酸化部位 tion Sites	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法およびデータベース
1	6471113CD1	793	S108 S128 S216 S272 S412 S435 S443 S485 S528 S582 S611 S616 S665 S8 T137 T148 T181 T502 T517 T606 T74	N246 N381 N382 N580 N604 N93	シグナルペプチド: M1-G35 シグナルペプチド: M1-G35 膜貫通ドメイン: I623-F641 免疫グロブリン ドメイン: G55-A119, G153-A211, H252-A309, C344-A402, フィブロネクチンタイプIII ドメイン: P424-S510, A521-T606 受容体チロシンキナーゼ BL00790: E439-A464, D487-T517 Score/strength=0.68 推定神経細胞接着分子 PUNC PD123980: N604-A783 塩基性線維芽細胞成長因子 受容体 1 DM01287: A39752|1-814: P244-E421, L16-V118, Q134-V210 P16092|1-821: P244-A420, W17-G124 A60350|1-800: V126-G374, G344-E421 免疫グロブリン DM00001|P43146|328-410: P329-Q410	SPSCAN HMMER HMMER HMMER-PFAM HMMER-PFAM BLIMPS-BLOCKS BLAST-PRODOM BLAST-DOMO BLAST-DOMO
2	7701502CD1	475	S220 S318 S85 T143 T203 T229 T322 T357 T393 T67 Y321	N100 N355 N364	シグナル切断: M1-A16 膜貫通ドメイン: A457-M475 免疫グロブリン ドメイン: G157-V217, G269-A325, G373-A427 細胞 前駆体 糖タンパク質 膜貫通 免疫グロブリンフォールド 接着 選択的スプライシング: PD005607: L2-R231	SPSCAN HMMER HMMER-PFAM BLAST-PRODOM

【表４】

表3－2

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸 残基	潜在的 リン酸化部位 tion Sites	潜在的 グリコシル 化部位	シグネチャ配列、 ドメインおよびモチーフ	分析 方法および データベース
2					ミエリン;シュワン;シアロアドヘジン;形成 DM03744|P20138|1-142:M1-T143	BLAST-DOMO
					免疫グロブリンおよび主要組織適合性複合体タンパク質 シグネチャ F423-H429	MOTIFS
3	5727379CD1	578	S109 S169 S315 S364 S477 T132 T307 T359 T378 T418 T425 T471 Y234 Y553	N130 N188 N250 N260 N299 N38 N463 N494 N499	膜貫通ドメイン: M519-W538	HMMER
					ロイシンリッチリピート: S202-F225, A318-Q340, E58-P81, L82-F105, N130-F153, S178-R201, N106-S129, N154-L177, N250-E273, F294-R317	HMMER-PFAM
					免疫グロブリンドメイン: G397-A457	HMMER-PFAM
					ロイシンリッチリピート N 末端ドメイン; G27-P56	HMMER-PFAM
4	5831801CD1	298	S125 S240 S266 S62 T156 T60 T90	N104 N138 N167 N172 N189 N35	シグナルペプチド: M1-G33	HMMER
					膜貫通ドメイン: I243-W263	HMMER
					免疫グロブリンドメイン: G48-Y120, S161-V219	HMMER-PFAM
					抗原 前駆体 シグナル 免疫グロブリンフォールド 糖タンパク質 T 細胞表面 膜貫通: PDC10953: I43-V231	BLAST-PRODOM
5	7195090CD1	2066	S1036 S1063 S109 S1104 S1173 S1267 S1327 S1368 S1422 S1440 S1501 S1523 S1590 S1620 S1695 S1779 S1788 S192 S1922 S1941 S1952 S1971	N1156 N1207 N145 N1528 N1622 N1641 N1693 N175 N1767 N1914 N1955 N424 N437 N446 N529 N553	膜貫通ドメイン: W1882-F1901	HMMER
					フィブロネクチンタイプIIIドメイン: P542-S629, P642-S731, P743-S834, P846-S932, P944-S1034, P1047-S1138, P1150-S1239, P1252-L1338, P1351-S1439, P1452-S1563, P1575-S1664, P1676-Q1765, P1777-S1864	HMMER-PFAM

【表５】

表3－3

SEQ ID NO:	Incyte ポリペプチド ID	アミノ酸残基	潜在的リン酸化部位 化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析 方法およびデータベース
5			S1989 S1998 S2005 S2029 S243 S398 S478 S652 S66 S853 S861 S865 S956 T1055 T1058 T1112 T1247 T126 T1300 T1307 T1312 T1345 T1365 T1447 T1507 T1534 T1624 T1917 T1923 T2003 T202 T222 T227 T25 T345 T357 T364 T386 T495 T632 T697 T713 T816 T827 T841 T900 T994	N656 N695 N756 N889 N898	免疫グロブリンドメイン: I30-V45, P93-Y134, S182-A238, E275-A334, G368-A429, G462-I523, C824-E903	HMMER-PFAM
					フィブロネクチンタイプIII 1ピート: T860-P869, N672-Y682, V595-Y613, Y1848-E1862	BLIMPS-PRINTS
					塩基性線維芽細胞成長因子 受容体 1 DM01287|A39752|1-814: E267-S398, P324-V444, L153-F245	BLAST-DOMO
					免疫グロブリン: DM00001|P20241|428-516: P354-V444	BLAST-DOMO
					フィブロネクチンタイプIII1ピート: DM00007|S50893|585-685: G624-Y722	BLAST-DOMO
6	7473826CD1	333	S310 S69 S89 T101 T164 T230 T316 T325 T35 T5	N125 N160 N256 N261 N33 N65	シグナルペプチド: M1-A60	SPSCAN
					IPT/TIG ドメイン(プレキシンと転写因子共有の免疫グロブリン様フォールド): P131-A211, P216-P298	HMMER-PFAM
					受容体キナーゼ チロシンタンパク質 前駆体 チロシンタンパク質 シグナル 肝細胞 成長 ATP 結合 PD003981: S118-N238, T2-P76, P223-V269 e-value=1.3e-10	BLAST-PRODOM

【表６】

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	選択された断片	配列断片	5' 位置	3' 位置
7	6471113CB1	2700	1-627, 1-139, 1711-2116, 410-1502, 1859-2134	7184821H1 (BONRFEC01)	2152	2700
				7744009H1 (ADRETUE04)	104	686
				5048190F8 (PLACFER01)	512	1040
				GNN.g8389522_000026_002.edit	410	2382
				1449087H1 (PLACNOT02)	1614	1857
				CNN.g8389522_000027_002.edit	1	409
				6471113F8 (PLACFEB01)	977	1389
				5048190R8 (PLACFER01)	796	1296
				7155347GV1	1966	2464
				2414883H1 (HNT3AZT01)	1395	1617
8	7701502CB1	2321	1915-1938, 1242-1385, 1-1001, 1988-2063, 1453-1503	7158364BV1	1953	2321
				7158374GV1	1348	2090
				7157114ZV1	1302	1916
				7168805H1 (MCLRNOC01)	537	1116
				4991695T6 (LIVRTUT11)	1	621
				1709963F6 (PROSNOT16)	674	1333
9	5727379CB1	2112	2052-2112, 1-456, 104?-1128, 864-1119	FL5727379_g8176790_000004_g3157804	1	1734
				7065822H1 (BRATNOR01)	1516	2112
10	5831801CB1	1419	1-262, 1 452, 1123-1364, 1368-1419	6884626H1 (BRAHTDR03)	873	1419
				7091169R8 (BRAUTDR03)	1	802
				6891983H1 (BRATTDR03)	800	1381
				7091169F8 (BRAUTDR03)	652	1301

【表７】

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	選択された断片	配列断片	5' 位置	3' 位置
11	7195090CB1	6374	1007-2395, 2744-6374, 1-144, 490-6119	7674279H2 (NOSETUE01)	3281	3386
				7674694J1 (NOSETUE01)	4711	5268
				g2993996	178	489
				7262823H1 (UTRETMC01)	2105	2736
				7195090H1 (LUNGFER04)	1908	2464
				7118549IV1	5722	6374
				7634914H1 (SINTDIE01)	2677	3235
				5600269SJ1	3863	4838
				7198117R8 (LUNGFER04)	243	1029
				7634914J1 (SINTDIE01)	3304	3966
				GNN.g6042167_008.edit	1	335
				7208025ID1	5056	5809
				7198133F8 (LUNGFER04)	947	1605
				6879120F8 (LNODNOR03)	1235	2011
				GBI.g10334962_000008_000011.edit	2633	3445
12	7473826CB1	1350	1250-1350, 1-341, 1-982, 760-1063	5001304F8 (PROSTUT21)	817	1350
				FL7473826_g7382420_000006_g6006555	1	1002

【表８】

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
7	6471113CB1	HNT2AZS07
8	7701502CB1	EOSINOT02
9	5727379CB1	UTRSNOT12
10	5831801CB1	BRAUTDR03
11	7195090CB1	LUNGFER04
12	7473826CB1	PROSTUT21

【表９】

表６－１

ライブラリ	ベクター	ライブラリの説明
BRAUTDR03	PCDNA2.1	このランダムプライムされたライブラリは胆管癌で死亡した 55 才の白人女性から摘出され、プールされていた淡蒼球および無名質組織から単離された RNA を用いて作製した。病理は弓隆部にわたる軽度の髄膜線維形成があり、帯状皮質白質および視床に軸索球状体が散在しており、内嗅野および中脳水道灰白質領域にいくぶん神経原線維濃縮体の散在があった。関連腫瘍組織の病理は、残留性または再発性腫瘍のある高分化型の肝臓胆管癌を示した。患者の病歴は胆管癌、術後パッド・キアーリ症候群、胆管腹水症、水胸症、脱水症、栄養不良、乏尿および急性腎不全が含まれる。前に受けた手術には胆嚢切除術および肝臓の85％の摘除がある。
EOSINOT02	PSPORT1	ライブラリはアレルギー性喘息患者から得たプールされていた好酸球から単離された RNA を用いて作成した。
HNT2AZS07	PSPORT1	このサブトラクションライブラリは 0.35 ミクロモルの AZ で3日間処理された hNT2 細胞系（コミットした神経細胞前駆体の特性を示すヒト奇形癌腫に由来する）から単離された RNA から作製した。サブトラクション用ハイブリダイゼーションプローブは未処理の hNT2 細胞から同様に作製されたライブラリから由来した。AZ 処理ライブラリからの 308 方のクローンは、未処理のライブラリからの 304 万のクローンとサブトラクションハイブリダイゼーションを3回行ったサブトラクションハイブリダイゼーションの条件は Swaroop ら, NAR(1991)19:1954 および Bonaldo らの Genome Research 6 (1996):791 の方法に基づいたものである。
LUNGFER04	PCDNA2.1	このランダムにプライムされたライブラリは胎児死亡の白人男子胎児から摘除した肺組織から単離された RNA を用いて作成した。

【表１０】

表6－2

ライブラリ	ベクター	ライブラリの説明
PROSTUT21	pINCY	ライブラリは、61才白人男性の根治的前立腺摘出術および部分的リンパ節切除術時に摘出された右前立腺腫瘍組織から単離されたRNAを用いて作製された。病理検査では、Gleasonグレード3+3の腺癌が右側中心および周辺にわたり主要な塊を形成していた。腫瘍は右側中後方嚢胞を侵襲していたが、それを超えていなかった。外科的縁は右尖部および右膀胱底部を含めて陽性であった。患者は硬結、前立腺の肥厚および前立腺特異抗原の上昇を示した。患者の病歴には腎不全、骨関節炎、左腎動脈狭窄、良性頂血圧、血小板減少症、高脂血症、ニコチン中毒およびアルコール中毒寛解期、およびC型肝炎(キャリア)が含まれる。前回の手術には股関節全置換術が含まれる。患者の服用薬剤にはアスピリンとインテラルがある。家族の病歴としては、母親と祖父母に良性の高血圧があった。
UTRSNOT12	pINCY	ライブラリは、掻爬および腟式子宮摘出術を行った白人女性(41才)から取り除かれた子宮筋層組織から単離したRNAを用いて作製した。子宮内膜は分泌性であり、子宮内膜ポリープの断片が含まれていた。子宮頸管内に良性の頸管内外粘膜が確認された。関連する腫瘍組織の病理学検査では、子宮平滑筋腫が見られた。患者の病歴には腹部ヘルニアおよび良性の卵巣腫瘍がある。

【表１１】

表7－1

プログラム	説明	引用文献	パラメータ閾値
ABI FACTURA	核酸配列においてベクター配列を除去して、不定の塩基をマスクするプログラム	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	Fast Data Finderは、アミノ酸または核酸配列の比較、注釈付けに有用である。	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	不一致<50%
ABI AutoAssembler	核酸配列を構築するプログラム。	Applied Biosystems, Foster City, CA.	
BLAST	Basic Local Alignment Search Tool はアミノ酸および核酸配列の配列類似性検索に有用であり、BLASTはblastp、blastn、blastx,、tblastnおよびtblastxの５つの機能がある。	Altschul, S.F. 他 (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. 他 (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: 確率値=1.0E-8以下 完全長配列：確率値1.0E-10 以下
FASTA	問合せ配列と同種の配列群配との類似性を検索するPearson およびLipman アルゴリズム。FASTAには最小5つの機能(fasta、tfasta、fastx, tfastxおよびssearch)がある。	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; Smith, T.F. 及び M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E 値=1.06E-6 構築された*ESTs*: fasta I同一性=95%以上 一致した長さ=200塩基以上 fastx E 値=1.0E-8 以下 *完全長配列:* fastx スコア=100以上
BLIMPS	配列をBLOCKS、PRINTS、DOMO、PRODOMおよびPFAM データベースの配列と対応させて遺伝子ファミリ、配列相同性および構造的フィンガープリント領域を検索するBLocks IMProved サーチャー。	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; 及びAttwood, T.K. 他 (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	確率値= 1.0E-3以下
HMMER	PFAMのようなタンパク質ファミリコンセンサス配列の隠れマルコフモデル(HMM)に基づいたデータベースに対して問合せ配列を検索するアルゴリズム。	Krogh, A. 他(1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. 他 (1988) Nucleic Acids Res. 26:320-322; Durbin, R. 他 (1998) Our World View, in a Nutshell, Cambridge Univ. Press, 1-350頁.	*PFAM*ヒット:確率値=1.0E-3 以下 シグナルペプチドヒット：スコア=０以上

表7－2

プログラム	説明	引用文献	パラメータ閾値
ProfileScan	Prositeで定義された配列パターンと一致するタンパク質配列の構造的モチーフおよび配列モチーフを検索するアルゴリズム。	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.	標準化された質のスコア≥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塩基読出しアルゴリズム。	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なインプリメンテーションに基づくSWATやCrossMatchを含むPhils Revised Assembly プログラムで、配列相同性の検索やDNA配列の構築に有用である。	Smith, T.F. 及び M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. 及び M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	スコア=120以上 一致した長さ=56以上
Consed	Phrap で構築したもの表示、編集用グラフィカルツール。	Gordon, D. 他 (1998) Genome Res. 8:195-202.	
SPScan	タンパク質配列をスキャンして、分泌シグナルペプチドの存在を調べる重み付けマトリックス解析プログラム。	Nielson, 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	加重マトリックスを用いて蛋白配列での膜貫通セグメントの描写および配向を決定するプログラム。	Persson, B. およびP. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. 及びP. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	隠れMarkov モデル(HMM) を使って蛋白配列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, eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, 175-182頁.	
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頁, 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
23 May 2002 (23.05.2002)

PCT

(10) International Publication Number
WO 02/40671 A2

(51) International Patent Classification⁷: C12N 15/12, C07K 14/705, 16/28, A01K 67/027, A61K 38/16, C12N 5/10, G01N 33/50, 33/53

(21) International Application Number: PCT/US01/44974

(22) International Filing Date:
15 November 2001 (15.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/249,643 16 November 2000 (16.11.2000) US

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

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandero, CA 94577 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **ELLIOTT, Vicki, S.** [US/US]; 5770 Polton Place Way, San Jose, CA 95121 (US). **THANGAVELU, Kavitha** [IN/US]; 1950 Montecito Avenue, #23, Mountain View, CA 94043 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **LU, Yan** [CN/US]; 3885 Corina Way, Palo Alto, CA 94303 (US). **LO, Terrence, B.** [CA/US]; 1451 Beach Park Boulevard, #115, Foster City, CA 94404 (US). **GURURAJAN, Rajagopal** [IN/US]; 5591 Dent Avenue, San Jose, CA 95118 (US). **GANDHI, Ameena, R.** [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). **ARVIZU, Chandra** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (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, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, 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:
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 02/40671 A2

(54) Title: IMMUNOGLOBULIN SUPERFAMILY PROTEINS

(57) Abstract: The invention provides human immunoglobulin superfamily proteins (IGSFP) and polynucleotides which identify and encode IGSFP. 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 IGSFP.

IMMUNOGLOBULIN SUPERFAMILY PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of immunoglobulin superfamily proteins and to the use of these sequences in the diagnosis, treatment, and prevention of immune system disorders, neurological disorders, developmental disorders, muscle disorders, and cell proliferative disorders, including cancer, 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

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 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-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. 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.

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

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 glycophosphatidylinositol 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-

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

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. 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 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, 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 neurotropin 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, 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. (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 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

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 β2 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; Paietta, 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

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

Leucine-Rich Repeat Proteins

Leucine-rich repeat proteins (LRRPs) are involved in protein-protein interactions. LRRPs 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 LRRPs 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 XCoe2, 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 Biochem. Sci. 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).

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 disorders, neurological disorders, developmental disorders, muscle disorders, and cell proliferative disorders, including cancer, 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-5," and "IGSFP-6." 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:1-6, 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, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. In one alternative, the

invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-6.

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:1-6, 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, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-6. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:7-12.

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:1-6, 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, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, 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, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. 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.

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:1-6, 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, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, 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:7-12, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, 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:7-12, 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 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.

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:7-12, 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:7-12, 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:1-6, 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, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. 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:1-6, 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, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. 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 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:1-6, 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, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group

consisting of SEQ ID NO:1-6. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional IGSFP, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, 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, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, 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, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting

altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, 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:7-12, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, 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:7-12, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

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

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability 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 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

"IGSFP" 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 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. 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'-NH_2), 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.)

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.

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 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 IGSFP or fragments of IGSFP 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
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

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

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

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

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

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

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

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

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

A fragment of SEQ ID NO:7-12 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:7-12, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:7-12 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:7-12 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:7-12 and the region of SEQ ID NO:7-12 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 fragment of SEQ ID NO:1-6 is encoded by a fragment of SEQ ID NO:7-12. A fragment of SEQ ID NO:1-6 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-6. For example, a fragment of SEQ ID NO:1-6 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-6. The precise length of a fragment of SEQ ID NO:1-6 and the region of SEQ ID NO:1-6 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

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

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

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

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

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2

Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

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

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

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

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e

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

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

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

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

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

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

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

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

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

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

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide

sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

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

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

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.

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

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

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

"Post-translational modification" of an 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 isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

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

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

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to

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

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

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use 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, 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.

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

transformed cells which express the inserted DNA or RNA for limited periods of time.

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

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

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

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

THE INVENTION

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

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

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank 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 polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7

shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are immunoglobulin superfamily proteins. For example, SEQ ID NO:1 is 87% identical to murine punc (putative neural cell adhesion protein) (GenBank ID g3068592) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains immunoglobulin and fibronectin type III 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 BLIMPS analysis provides further corroborative evidence that SEQ ID NO:1 is a neural cell adhesion protein. SEQ ID NO:2-6 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-6 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:7-12 or that distinguish between SEQ ID NO:7-12 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 7184821H1 is the identification number of an Incyte cDNA sequence, and BONRFEC01 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71553470V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g2993996) which contributed to the assembly of the full

length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL_*XXXXXX*_N_1_N_2_*YYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is the identification number of a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* 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.*, g*BBBBB*).

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

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).

INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

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

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

The invention also encompasses 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:7-12, which encodes IGSFP. The polynucleotide sequences of SEQ ID NO:7-12, 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 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:7-12 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:7-12. 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 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 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. 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 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 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 its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding 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 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 NO:7-12 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols 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, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo

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

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

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

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

In another embodiment, sequences encoding 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 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. 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.)

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 promotors, 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 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 be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

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

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 with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA

90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding IGSFP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding IGSFP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding IGSFP into the vector's multiple cloning site disrupts the *lac*Z gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. 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 containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of 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.

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 herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β glucuronidase and its substrate β-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to

quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding 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 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 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 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, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding 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 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 case 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 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 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 (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 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, ^{35}S-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 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 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 eosinophils, lung tissue, prostate tumor tissue, uterine myometrial tissue, and brain tissues such as globus pallidus and substantia innominata. Therefore, IGSFP appears to play a role in immune system disorders, neurological disorders, developmental disorders, muscle disorders, and cell proliferative disorders, including cancer. 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 (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, 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, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease,

stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system 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, 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, 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 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 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 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 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, 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 prevent a disorder associated with increased expression or activity of IGSFP. Examples of such disorders include, but are not limited to, those immune system disorders, neurological disorders, developmental disorders, muscle disorders, and cell proliferative disorders, including cancer, 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 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.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, 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 technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce 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 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, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

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

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 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_a 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_a ranging from about 10^9 to 10^{13} L/mole are preferred for use in immunoassays in which the IGSFP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of 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 gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding 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) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated

cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in 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 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. Récipon (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 (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 (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 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 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 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 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 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. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding 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.

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 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 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 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 (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

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

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. 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.

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

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-

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

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising 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).

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

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

The exact dosage will be determined by the practitioner, in light of factors related to the

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

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

DIAGNOSTICS

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

A variety of protocols for measuring 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 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 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 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.

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:7-12 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 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 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 immune system disorder such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia 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, 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, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's

syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system 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, 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, 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 neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham'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 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 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 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 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 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.

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 subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the

patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

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

Additional diagnostic uses for oligonucleotides designed from the sequences encoding 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 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 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 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 fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the

high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of 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. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

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

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.

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. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The

resultant transcript image would provide a profile of gene activity.

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

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

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

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

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

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

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

degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

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

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

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

In another embodiment of the invention, nucleic acid sequences encoding IGSFP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop

genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

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

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, 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 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 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.

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.

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.

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

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/249,645, are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

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

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using

the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., 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), pIGEN (Incyte Genomics, Palo Alto CA), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

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

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

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared

using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

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

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

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, 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:7-12. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative immunoglobulin superfamily proteins were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of 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

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 gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in

Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The 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:7-12 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:7-12 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

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

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

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 II). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided

by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding 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

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

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

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

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

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

digested with 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% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

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

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:7-12 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{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 hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

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

X. Microarrays

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

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

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first

strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse 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 the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. 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 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 μl of the array element DNA, at an average concentration of 100 ng/μl, 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 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 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

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

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

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

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital

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

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

XI. 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.

XII. 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 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 XVI and XVII, where applicable.

XIII. 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 include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light

scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of 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 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 analysis or microarray techniques.

XIV. 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 rabbits 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 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.

XV. 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 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 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.

XVI. Identification of Molecules Which Interact with IGSFP

IGSFP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules 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 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. Patent No. 6,057,101).

XVII. 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, pp. 113-115). 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 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. 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 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 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.

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

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
6471113	1	6471113CD2	7	6471113CB1
7701502	2	7701502CD1	8	7701502CB1
5727379	3	5727379CD1	9	5727379CB1
5831801	4	5831801CD1	10	5831801CB1
7195090	5	7195090CD1	11	7195090CB1
7473826	6	7473826CD1	12	7473826CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
1	6471113CD1	g3068592	0.0	Punc [Mus musculus] (Saltzman, J.M. (1998) Mech. Dev. 71:201-204)
2	7701502CD1	g6746356	8.8e-74	sialic acid-binding immunoglobulin-like lectin-8 [Homo sapiens] (Floyd, H. et al. (2000) J. Biol. Chem. 275:861-866)
3	5737379CD1	g9837568	7.1e-43	neuronal leucine-rich repeat protein-3 [Rattus norvegicus]
4	5831801CD1	g312590	3.4e-15	biliary glycoprotein [Mus musculus] (McCuaig, K. et al. (1993) Gene 127:173-183)
5	7195090CD1	g7290075	0.0	sdk gene product [Drosophila melanogaster] (Adams, M.D. et al. (2000) Science 287:2185-2195)

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	6471113CD1	793	S108 S128 S216 S272 S413 S436 S443 S486 S528 S582 S611 S616 S665 S8 T137 T148 T181 T502 T517 T606 T74	N246 N361 N382 N580 N604 N93	Signal peptide: M1-G35 Signal peptide: M1-G35 Transmembrane domain: I673-F641 Immunoglobulin domains: G56-A119, G153-A211, H252-A309, G344-A402, Fibronectin type III domains: P424-S510, A521-N606 Receptor tyrosine kinase BL00790: E439-A464, D487-T517 Score/strength=0.68 PUTATIVE NEURONAL CELL ADHESION MOLECULE PUNC PD123380: N604-A783 BASIC FIBROBLAST GROWTH FACTOR RECEPTOR 1 PR01287: A39752|1-814: P244-R422, L16-V118, Q134-V210 P16092|1-821: P244-A420, W17-G124 A60350|1-800: V126-G374, G344-S421 IMMUNOGLOBULIN EX00001|P43146|328-410: P329-Q410	SEQUENCE HMMER HMMER HMMER-PFAM HMMER-PFAM BLIMPS-BLOCKS BLAST-PRODOM BLAST-DOMO BLAST-DOMO
2	7701502CD1	475	S220 S318 S85 T143 T203 T229 T322 T357 T393 T67 Y321	N100 N355 N364	Signal cleavage: M1-A16 Transmembrane domain: A457-M475 Immunoglobulin domain: G157-V217, G269-A325, G373-A427 CELL RECOGNITION GLYCOPROTEIN TRANSMEMBRANE IMMUNOGLOBULIN FOLD ADHESION ALTERNATIVE SPLICING: PD005907: L2-A231	SPSCAN HMMER HMMER-PFAM BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polynucleotide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2					MEMBRN; SCHWANN; STARGARDNSTN; FORM DM037444|P20138|1-142:M1-T143	BLAST-DOMO
					Immunoglobulins and major histocompatibility complex proteins signature F423-H429	MOTIFS
3	5727379CD1	578	S109 S169 S315 S364 S477 T132 T307 T359 T378 T418 T435 T471 Y234 Y553	N130 N188 N250 N260 N299 N38 N463 N494 N499	Transmembrane domain: M519-W538	HMMER
					Leucine Rich Repeat: S203-P225, A218-Q340, E56-P81, L82-E105, N130-H153, S178-R201, N106-S129, K154-L177, N250-E273, F294-R317	HMMER-PFAM
					Immunoglobulin domain: G397-A457	HMMER-PFAM
					Leucine rich repeat N-terminal domain: G27-P56	HMMER-PFAM
4	5831801CD1	298	S125 S240 S266 S62 T156 Y60 T90	N104 N138 N167 N172 N189 N35	Signal peptide: M1-G33	HMMER
					Transmembrane domain: I243-W263	HMMER
					Immunoglobulin domains: G48-Y120, S161-V219	HMMER-PFAM
					ANTIGEN PRECURSOR SIGNAL IMMUNOGLOBULIN FOLD GLYCOPROTEIN T CELL SURFACE TRANSMEMBRANE: BD010953: I43-V231	BLAST-PRODOM
5	7195090CD1	2066	S1036 S1063 S109 S1104 S1173 S1287 S1327 S1368 S1422 S1440 S1501 S1523 S1590 S1620 S1695 S1779 S1788 S192 S1922 S1941 S1952 S1971	N1156 N1207 N145 N1520 N1622 N1641 N1693 N175 N1767 N1914 N1955 N424 N437 N446 N529 N553	Transmembrane domain: W1882-F1901	HMMER
					Fibronectin type III domains: E542-S629, E662-S731, P743-S834, E846-S932, F944-S1034, P1047-S1138, P1150-S1239, P1252-L1338, P1351-S1439, P1452-S1563, P1575-S1664, P1676-Q1765, P1777-S1864	HMMER-PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5			S1989 S1998 S2005 S2029 S243 S398 S478 S652 S66 S853 S862 S865 S956 T1055 T1058 T1113 T1247 T126 T1300 T1307 T1312 T1345 T1365 T1447 T1507 T1534 T1624 T1917 T1923 T2003 T202 T222 T227 T25 T345 T357 T364 T386 T495 T632 T697 T713 T816 T827 T841 T900 T994	N656 N695 N756 N889 N898	Immunoglobulin domains: I30-Y45 P93-Y134, S182-A236, E275-A334, G368-A429, G462-I523, C824-E903. Fibronectin type III repeats PR00014: T860-P869, M672-Y682, V535-Y613, Y1848-E1862 BASIC FIBROBLAST GROWTH FACTOR RECEPTOR 1 BM01287|A39752|1-814: S267-S398, P324-V444, G153-P245 IMMUNOGLOBULIN: DM00001|P20342|428-516: P354-V444 FIBRONECTIN TYPE III REPEAT: DM00007|S50893|585-685: G624-Y722	HMMER-PFAM BLIMPS-PRINTS BLAST-DOMO BLAST-DOMO BLAST-DOMO
6	7479836CD1	333	S310 S69 S89 T101 T164 T230 T316 T325 T35 T5	N125 N160 N256 N261 N33 N65	Signal peptide: M1-A60 IPT/TIG domains (immunoglobulin- like fold shared by plexins and transcription factors): P131-A211, P216-P298 RECEPTOR KINASE TYROSINE PROTEIN PRECURSOR TYROSINEPROTEIN SIGNAL HEPATOCYTE GROWTH ATPBINDING BD003981: S118-M238, T2-P76, P223-V269 e-value=1.3e-10	SPSCAN HMMER-PFAM BLAST-PRODOM

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
7	6471113CB1	2700	1-627, 1-139, 1711-2116, 410-1502, 1859-2134	7134821H1 (BONNFEC01)	2152	2700
				7744009H1 (ADRETUE04)	104	686
				5048190F8 (PLACFEB01)	512	1040
				GNN.gb389522_000026_002.edit	410	2382
				1449007H1 (PLACNOT02)	1614	1857
				GNN.gb389522_000027_002.edit	1	409
				6471113F8 (PLACFEB01)	977	2389
				5048190H8 (PLACFEB01)	796	1296
				7155347OV1	1966	2464
				2414883H1 (HNT3AZT01)	1395	1617
8	7701502CB1	2321	195-1038, 1242-1385, 1-1001, 1968-2063, 1459-1503	7158364BV1	1953	2321
				7158174OV1	1348	2090
				7157114ZV1	1302	1916
				7168805H1 (MCLRNOC01)	537	1116
				4991695T6 (LIVTUT11)	1	621
				1709963F6 (PROSNOT16)	674	1333
9	5727379CB1	2112	2052-2112, 1-456, 1044-1128, 864-1119	GI_5727379_g8176790_000004_g315_7804	1	1734
				7065822H1 (BRATNOR01)	1516	2112
10	5831801CB1	1419	1-262, 2-452, 1123-1364, 1368-1419	6884626H1 (BRAETDR03)	873	1419
				7091169R8 (BRAUTDR03)	1	802
				6891983H1 (BRAITDR03)	800	1381
				7091169F8 (BRAUTDR03)	652	1301

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
11	7195090CB1	6374	1007-2395, 2744-6374, 1-144, 490-6119	7674279H2 (NOSEMUB01)	3281	3386
				7674694J1 (NOSEMUB01)	4711	5268
				g2993996	178	409
				7262823F1 (UTRBMBC01)	2105	2736
				7195090R1 (LUNGFER04)	1908	2464
				7118549IV1	3722	6374
				7634914B1 (SINFDIE01)	2677	3235
				5600269S1	3863	4838
				7159841TR8 (LUNGFER04)	243	1029
				7634914J1 (SINFDIE01)	3304	3966
				GNN.g60342167_008.edit	1	335
				7208025ID1	5056	5809
				7198133F8 (LUNGFER04)	947	1605
				6879120F8 (LNCDNOR03)	1235	2011
				GBI.g40334962_000008_000011.edit	2633	3445
12	7473826CB1	1350	1250-1350, 1-341, 1-982, 760-1063	5001304F8 (PROSTUT21)	817	1350
				FL7473826_gi7382420_000006_g600 6555	1	1002

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
7	6471113CB1	HWP2AZS07
8	7705902CB1	BOSNOF02
9	5737379CB1	UTSANOB12
10	5831801CB1	BRAUNDR03
11	7195090CB1	JURGFER04
12	7473826CB1	PROSTUT21

Table 6

Library	Vector	Library Description
BRAUTDR03	PCDNA2.1	This random primed library was constructed using RNA isolated from pooled globus pallidus and substantia innominata tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BOSINOT02	PSPORT1	Library was constructed using RNA isolated from pooled eosinophils obtained from allergic asthmatic individuals.
NNT2AZS07	PSPORT1	This subtracted library was constructed from mRNA isolated from an hNT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor) treated for three days with 0.35 micromolar Ara. The hybridization probe for subtraction was derived from a similarly constructed library from untreated hNT2 cells. 3.03M clones from the Ara-treated library were subjected to three rounds of subtractive hybridization with 3.04M clones from the untreated library. Subtractive hybridization conditions were based on the methodologies of Swaroop et al. (NAR (1991) 19:1954) and Bonaldo et al. (Genome Research (1996) 6:791).
LUNGFER04	PCDNA2.1	This random primed library was constructed using RNA isolated from lung tissue removed from a Caucasian male fetus who died from fetal demise.

Table 6 (cont.)

Library	Vector	Library Description
PROSTUT21	pINCY	Library was constructed using RNA isolated from right prostate tumor tissue removed from a 61-year-old Caucasian male during a radical prostatectomy and regional lymph node excision. Pathology indicated adenocarcinoma, Gleason grade 3+3, forming a predominant mass involving the right side centrally and peripherally. The tumor invaded the right mid-posterior capsule but did not extend beyond it. The surgical margins were positive, including right apex and right bladder base. The patient presented with induration, hyperplasia of the prostate, and elevated prostate specific antigen. Patient history included renal failure, osteoarthritis, left renal artery stenosis, benign hypertension, thrombocytopenia, hyperlipidemia, tobacco and alcohol abuse in remission, and hepatitis C (carrier). Previous surgeries included a total hip replacement. Patient medications included aspirin and Indocin. Family history included benign hypertension in the mother and in the grandparent(s).
UTRSNOT12	pINCY	Library was constructed using RNA isolated from uterine myometrial tissue removed from a 41-year-old Caucasian female during a vaginal hysterectomy with dilation and curettage. The endometrium was secretory and contained fragments of endometrial polyps. Benign endo- and ectocervical mucosa were identified in the endocervix. Pathology for the associated tumor tissue indicated uterine leiomyoma. Patient history included ventral hernia and a benign ovarian neoplasm.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL 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.	*EST's:* 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 as least five functions: fasta, tfasta, fasty, tfastx, and ssearch	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl Math. 2:482-489.	*EST's:* fasta E value=1.06E-6 *Assembled ESTs:* fasta Identity= 95% or greater and Match length=200 bases or greater; fasta E value=1.0E-8 or less *Full Length sequences:* fasta score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) 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
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits:* Probability value= 1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality scores/GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils 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.	
TMHMM/HMMR	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Int. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

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:1-6,

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,

c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and

d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

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:7-12.

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

polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

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:7-12,

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:7-12,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

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 which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

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

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

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

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

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

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

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

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.

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.

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.

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:

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.

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.

30. A diagnostic test for a condition or disease associated with the expression of IGSFP 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.

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.

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.

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.

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:1-6, 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 binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

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

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:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

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 library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6 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 polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6 from 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) 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:1-6.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

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 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 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 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.

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.

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.

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 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 polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:7.

63. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:8.

64. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:9.

65. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:10.

66. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:11.

67. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:12.

<110> INCYTE GENOMICS, INC.
SAUGEN, Mariah R.
LU, Dyung Aina M.
YUE, Henry
ELLIOTT, Vicki S.
THANGAVELU, Kavitha
RAMKUMAR, Jayalaxmi
LU, Yan
LO, Terrence P.
GURURAJAN, Rajagopal
GANDHI, Ameena R.
ARVIZU, Chandra
YAO, Monique G.

<120> IMMUNOGLOBULIN SUPERFAMILY PROTEINS

<130> PF-0841 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/249,645
<151> 2000-11-16

<160> 12
<170> PERL Program

<210> 1
<211> 793
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6471213CD1

<400> 1

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

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

```
                560                 565                 570
Leu Leu Pro Gly Thr Val Ser Ser Tyr Asn Leu Ser Gln Leu Asp
                575                 580                 585
Pro Thr Ala Val Tyr Glu Val Lys Leu Leu Ala Tyr Asn Gln His
                590                 595                 600
Gly Asp Gly Asn Ala Thr Val Arg Phe Val Ser Leu Arg Gly Ala
                605                 610                 615
Ser Glu Arg Thr Gly Ile Val Ile Gly Ile His Ile Gly Val Thr
                620                 625                 630
Cys Ile Ile Phe Cys Val Leu Phe Leu Leu Phe Gly Gln Arg Gly
                635                 640                 645
Arg Val Leu Leu Cys Lys Asp Val Glu Asn Gln Leu Ser Pro Pro
                650                 655                 660
Gln Gly Pro Arg Ser Gln Arg Asp Pro Gly Ile Leu Ala Leu Asn
                665                 670                 675
Gly Ala Arg Arg Gly Gln Arg Gly Gln Leu Gly Arg Asp Glu Lys
                680                 685                 690
Arg Val Asp Met Lys Glu Leu Glu Gln Leu Phe Pro Pro Ala Ser
                695                 700                 705
Ala Ala Gly Gln Pro Asp Pro Arg Pro Thr Asp Pro Ala Ala Pro
                710                 715                 720
Ala Pro Cys Glu Glu Thr Gln Leu Ser Leu Leu Pro Leu Gln Gly
                725                 730                 735
Cys Gly Leu Met Glu Gly Lys Thr Thr Glu Ala Lys Thr Thr Glu
                740                 745                 750
Ala Thr Ala Pro Cys Ala Gly Leu Ala Ala Ala Pro Pro Pro Pro
                755                 760                 765
Asp Gly Gly Pro Gly Leu Leu Ser Glu Gly Gln Ala Ser Arg Pro
                770                 775                 780
Ala Ala Ala Arg Val Thr Gln Pro Ala His Ser Glu Gln
                785                 790

<210> 2
<211> 475
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7701502CD1

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

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

<210> 3
<211> 578
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5727379CD1

<400> 3
Met Leu His Thr Ala Ile Ser Cys Trp Gln Pro Phe Leu Gly Leu
1 5 10 15
Ala Val Val Leu Ile Phe Met Gly Ser Thr Ile Gly Cys Pro Ala
20 25 30
Arg Cys Glu Cys Ser Ala Gln Asn Lys Ser Val Ser Cys His Arg
35 40 45
Arg Arg Leu Ile Ala Ile Pro Glu Gly Ile Pro Ile Glu Thr Lys
50 55 60
Ile Leu Asp Leu Ser Lys Asn Arg Leu Lys Ser Val Asn Pro Glu
65 70 75
Glu Phe Ile Ser Tyr Pro Leu Leu Glu Glu Ile Asp Leu Ser Asp
80 85 90
Asn Ile Ile Ala Asn Val Glu Pro Gly Ala Phe Asn Asn Leu Phe
95 100 105
Asn Leu Arg Ser Leu Arg Leu Lys Gly Asn Arg Leu Lys Leu Val
110 115 120
Pro Leu Gly Val Phe Thr Gly Leu Ser Asn Leu Thr Lys Leu Asp
125 130 135
Ile Ser Glu Asn Lys Ile Val Ile Leu Leu Asp Tyr Met Phe Gln
140 145 150
Asp Leu His Asn Leu Lys Ser Leu Glu Val Gly Asp Asn Asp Leu
155 160 165
Val Tyr Ile Ser His Arg Ala Phe Ser Gly Leu Leu Ser Leu Glu
170 175 180
Gln Leu Thr Leu Glu Lys Cys Asn Leu Thr Ala Val Pro Thr Glu
185 190 195
Ala Leu Ser His Leu Arg Ser Leu Ile Ser Leu His Leu Lys His
200 205 210
Leu Asn Ile Asn Asn Met Pro Val Tyr Ala Phe Lys Arg Leu Phe
215 220 225
His Leu Lys His Leu Glu Ile Asp Tyr Trp Pro Leu Leu Asp Met
230 235 240
Met Pro Ala Asn Ser Leu Tyr Gly Leu Asn Leu Thr Ser Leu Ser
245 250 255
Val Thr Asn Thr Asn Leu Ser Thr Val Pro Phe Leu Ala Phe Lys
260 265 270
His Leu Glu Leu His Ile Val Gly Ala Gln Leu Arg Thr Ile Glu
275 280 285
Pro His Ser Phe Gln Gly Leu Arg Phe Leu Arg Val Leu Asn Val
290 295 300
Ser Gln Asn Leu Leu Glu Thr Leu Glu Glu Asn Val Phe Ser Ser
305 310 315
Pro Arg Ala Leu Glu Val Leu Ser Ile Asn Asn Asn Pro Leu Ala
320 325 330
Cys Asp Cys Arg Leu Leu Trp Ile Leu Gln Arg Gln Pro Thr Leu
335 340 345
Gln Phe Gly Gly Gln Gln Pro Met Cys Ala Gly Pro Asp Thr Ile
350 355 360
Arg Glu Arg Ser Phe Lys Asp Phe His Ser Thr Ala Leu Ser Phe
365 370 375
Tyr Phe Thr Cys Lys Lys Pro Lys Ile Arg Glu Lys Lys Leu Gln

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

<210> 4
<211> 298
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5831801CD1

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

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

<210> 5
<211> 2066
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7195090CD1

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

```
                170                 175                 180
Gly Ser Ser Glu Asn Thr Leu Glu Cys Ile Ala Ser Ala Arg Pro
                185                 190                 195
Val Glu Asp Leu Ser Val Thr Trp Lys Arg Asn Gly Val Arg Ile
                200                 205                 210
Thr Ser Gly Leu His Ser Phe Gly Arg Arg Leu Thr Ile Ser Asn
                215                 220                 225
Pro Thr Ser Ala Asp Thr Gly Pro Tyr Val Cys Glu Ala Ala Leu
                230                 235                 240
Pro Gly Ser Ala Phe Glu Pro Ala Arg Ala Thr Ala Phe Leu Phe
                245                 250                 255
Ile Ile Glu Pro Pro Tyr Phe Thr Ala Glu Pro Glu Ser Arg Ile
                260                 265                 270
Ser Ala Glu Val Glu Glu Thr Val Asp Ile Gly Cys Gln Ala Met
                275                 280                 285
Gly Val Pro Leu Pro Thr Leu Gln Trp Cys Lys Asp Ala Ile Ser
                290                 295                 300
Ile Ser Arg Leu Gln Asn Pro Arg Tyr Lys Val Leu Ala Ser Gly
                305                 310                 315
Gly Leu Arg Ile Gln Lys Leu Arg Pro Glu Asp Ser Gly Ile Phe
                320                 325                 330
Gln Cys Phe Ala Ser Asn Glu Gly Gly Glu Ile Gln Thr His Thr
                335                 340                 345
Tyr Leu Asp Val Thr Asn Ile Ala Pro Val Phe Thr Gln Arg Pro
                350                 355                 360
Val Asp Thr Thr Val Thr Asp Gly Met Thr Ala Ile Leu Arg Cys
                365                 370                 375
Glu Val Ser Gly Ala Pro Lys Pro Ala Ile Thr Trp Lys Arg Glu
                380                 385                 390
Asn His Ile Leu Ala Ser Gly Ser Val Arg Ile Pro Arg Phe Met
                395                 400                 405
Leu Leu Glu Ser Gly Gly Leu Gln Ile Asn Pro Val Phe Ile Gln
                410                 415                 420
Asp Ala Gly Asn Tyr Thr Cys Tyr Ala Ala Asn Thr Glu Gly Ser
                425                 430                 435
Leu Asn Ala Ser Ala Thr Leu Thr Val Trp Asn Arg Thr Ser Ile
                440                 445                 450
Val His Pro Pro Glu Asp His Val Val Ile Lys Gly Thr Thr Ala
                455                 460                 465
Thr Leu His Cys Gly Ala Thr His Asp Pro Arg Val Ser Leu Arg
                470                 475                 480
Tyr Val Trp Lys Lys Asp Asn Val Ala Leu Thr Pro Ser Ser Thr
                485                 490                 495
Ser Arg Ile Val Val Glu Lys Asp Gly Ser Leu Leu Ile Ser Gln
                500                 505                 510
Thr Trp Ser Gly Asp Ile Gly Asp Tyr Ser Cys Glu Ile Val Ser
                515                 520                 525
Glu Gly Gly Asn Asp Ser Arg Met Ala Arg Leu Glu Val Ile Glu
                530                 535                 540
Leu Pro His Ser Pro Gln Asn Leu Leu Val Ser Pro Asn Ser Ser
                545                 550                 555
His Ser His Ala Val Val Leu Ser Trp Val Arg Pro Phe Asp Gly
                560                 565                 570
Asn Ser Pro Ile Leu Tyr Tyr Ile Val Glu Leu Ser Glu Asn Asn
                575                 580                 585
Ser Pro Trp Lys Val His Leu Ser Asn Val Gly Pro Glu Met Thr
```

590 595 600
Gly Val Thr Val Ser Gly Leu Thr Pro Ala Arg Thr Tyr Gln Phe
605 610 615
Arg Val Cys Ala Val Asn Glu Val Gly Arg Gly Gln Tyr Ser Ala
620 625 630
Glu Thr Ser Arg Leu Met Leu Pro Glu Glu Pro Pro Ser Ala Pro
635 640 645
Pro Lys Asn Ile Val Ala Ser Gly Arg Thr Asn Gln Ser Ile Met
650 655 660
Val Gln Trp Gln Pro Pro Pro Glu Thr Glu His Asn Gly Val Leu
665 670 675
Arg Gly Tyr Ile Leu Arg Tyr Arg Leu Ala Gly Leu Pro Gly Glu
680 685 690
Tyr Gln Gln Arg Asn Ile Thr Ser Pro Glu Val Asn Tyr Cys Leu
695 700 705
Val Thr Asp Leu Ile Ile Trp Thr Gln Tyr Glu Ile Gln Val Ala
710 715 720
Ala Tyr Asn Gly Ala Gly Leu Gly Val Phe Ser Arg Ala Val Thr
725 730 735
Glu Tyr Thr Leu Gln Gly Val Pro Thr Ala Pro Pro Gln Asn Val
740 745 750
Gln Thr Glu Ala Val Asn Ser Thr Thr Ile Gln Phe Leu Trp Asn
755 760 765
Pro Pro Pro Gln Gln Phe Ile Asn Gly Ile Asn Gln Gly Tyr Lys
770 775 780
Leu Leu Ala Trp Pro Ala Asp Ala Pro Glu Ala Val Thr Val Val
785 790 795
Thr Ile Ala Pro Asp Phe His Gly Val His His Gly His Ile Thr
800 805 810
Asn Leu Lys Lys Phe Thr Ala Tyr Phe Thr Ser Val Leu Cys Phe
815 820 825
Thr Thr Pro Gly Asp Gly Pro Pro Ser Thr Pro Gln Leu Val Trp
830 835 840
Thr Gln Glu Asp Lys Pro Gly Ala Val Gly His Leu Ser Phe Thr
845 850 855
Glu Ile Leu Asp Thr Ser Leu Lys Val Ser Trp Gln Glu Pro Leu
860 865 870
Glu Lys Asn Gly Ile Ile Thr Gly Tyr Gln Ile Ser Trp Glu Val
875 880 885
Tyr Gly Arg Asn Asn Ser Arg Leu Thr His Thr Leu Asn Ser Thr
890 895 900
Thr His Glu Tyr Lys Ile Gln Gly Leu Ser Ser Leu Thr Thr Tyr
905 910 915
Thr Ile Asp Val Ala Ala Val Thr Ala Val Gly Thr Gly Leu Val
920 925 930
Thr Ser Ser Thr Ile Ser Ser Gly Val Pro Pro Asp Leu Pro Gly
935 940 945
Ala Pro Ser Asn Leu Val Ile Ser Asn Ile Ser Pro Arg Ser Ala
950 955 960
Thr Leu Gln Phe Arg Pro Gly Tyr Asp Gly Lys Thr Ser Ile Ser
965 970 975
Arg Trp Ile Val Glu Gly Glu Val Gly Ala Ile Gly Asp Glu Glu
980 985 990
Glu Trp Val Thr Leu Tyr Glu Glu Glu Asn Glu Pro Asp Ala Gln
995 1000 1005
Met Leu Glu Ile Pro Asn Leu Thr Pro Tyr Thr His Tyr Arg Phe

1010 1015 1020
Arg Met Lys Gln Val Asn Ile Val Gly Pro Ser Pro Tyr Ser Pro
1025 1030 1035
Ser Ser Arg Val Ile Gln Thr Leu Gln Ala Pro Pro Asp Val Asn
1040 1045 1050
Pro Thr Ser Val Thr Val Arg Thr Ala Ser Glu Thr Ser Leu Arg
1055 1060 1065
Leu Arg Trp Val Pro Leu Pro Asp Ser Gln Tyr Asn Gly Asn Pro
1070 1075 1080
Glu Ser Val Gly Tyr Arg Ile Lys Tyr Trp Arg Ser Asp Leu Gln
1085 1090 1095
Ser Ser Ala Val Ala Gln Val Val Ser Asp Arg Leu Glu Arg Glu
1100 1105 1110
Phe Thr Ile Glu Glu Leu Glu Glu Trp Met Glu Tyr Glu Leu Gln
1115 1120 1125
Met Gln Ala Phe Asn Ala Val Gly Ala Gly Pro Trp Ser Glu Val
1130 1135 1140
Val Arg Gly Arg Thr Arg Glu Ser Val Pro Ser Ala Ala Pro Glu
1145 1150 1155
Asn Val Ser Ala Glu Ala Val Ser Ser Thr Gln Ile Leu Leu Thr
1160 1165 1170
Trp Thr Ser Val Pro Glu Gln Asp Gln Asn Gly Leu Ile Leu Gly
1175 1180 1185
Tyr Lys Ile Leu Phe Arg Ala Lys Asp Leu Asp Pro Glu Pro Arg
1190 1195 1200
Ser His Ile Val Arg Gly Asn His Thr Gln Ser Ala Leu Leu Ala
1205 1210 1215
Gly Leu Arg Lys Phe Val Leu Tyr Glu Leu Gln Val Leu Ala Phe
1220 1225 1230
Thr Arg Ile Gly Asn Gly Val Pro Ser Thr Pro Leu Ile Leu Glu
1235 1240 1245
Arg Thr Lys Asp Asp Ala Pro Gly Pro Pro Val Arg Leu Val Phe
1250 1255 1260
Pro Glu Val Arg Leu Thr Ser Val Arg Ile Val Trp Gln Pro Pro
1265 1270 1275
Glu Glu Pro Asn Gly Ile Ile Leu Gly Tyr Gln Ile Ala Tyr Arg
1280 1285 1290
Leu Ala Ser Ser Pro His Thr Phe Thr Thr Val Glu Val Gly
1295 1300 1305
Ala Thr Val Arg Gln Phe Thr Ala Thr Asp Leu Ala Pro Glu Ser
1310 1315 1320
Ala Tyr Ile Phe Arg Leu Ser Ala Lys Thr Arg Gln Gly Trp Gly
1325 1330 1335
Glu Pro Leu Glu Ala Thr Val Ile Thr Thr Glu Lys Arg Glu Arg
1340 1345 1350
Pro Ala Pro Pro Arg Glu Leu Leu Val Pro Gln Ala Glu Val Thr
1355 1360 1365
Ala Arg Ser Leu Arg Leu Gln Trp Val Pro Gly Ser Asp Gly Ala
1370 1375 1380
Ser Pro Ile Arg Tyr Phe Thr Met Gln Val Arg Glu Leu Pro Arg
1385 1390 1395
Gly Glu Trp Gln Thr Tyr Ser Ser Ser Ile Ser His Glu Ala Thr
1400 1405 1410
Ala Cys Val Val Asp Arg Leu Arg Pro Phe Thr Ser Tyr Lys Leu
1415 1420 1425
Arg Leu Lys Ala Thr Asn Asp Ile Gly Asp Ser Asp Phe Ser Ser

1430 1435 1440
Glu Thr Glu Ala Val Thr Thr Leu Gln Asp Val Pro Gly Glu Pro
1445 1450 1455
Pro Gly Ser Val Ser Ala Thr Pro His Thr Thr Ser Ser Val Leu
1460 1465 1470
Ile Gln Trp Gln Pro Pro Arg Asp Glu Ser Leu Asn Gly Leu Leu
1475 1480 1485
Gln Gly Tyr Arg Ile Tyr Tyr Arg Glu Leu Glu Tyr Glu Ala Gly
1490 1495 1500
Ser Gly Thr Glu Ala Lys Thr Leu Lys Asn Pro Ile Ala Leu His
1505 1510 1515
Ala Glu Leu Thr Ala Gln Ser Ser Phe Lys Thr Val Asn Ser Ser
1520 1525 1530
Ser Thr Ser Thr Met Cys Glu Leu Thr His Leu Lys Lys Tyr Arg
1535 1540 1545
Arg Tyr Glu Val Ile Met Thr Ala Tyr Asn Ile Ile Gly Glu Ser
1550 1555 1560
Pro Ala Ser Ala Pro Val Glu Val Phe Val Gly Glu Ala Ala Pro
1565 1570 1575
Ala Met Ala Pro Gln Asn Val Gln Val Thr Pro Leu Thr Ala Ser
1580 1585 1590
Gln Leu Glu Val Met Trp Asp Pro Pro Pro Pro Glu Ser Gln Asn
1595 1600 1605
Gly Asn Ile Gln Gly Tyr Lys Ile Tyr Tyr Trp Glu Ala Asp Ser
1610 1615 1620
Gln Asn Glu Thr Glu Lys Met Lys Val Leu Phe Leu Pro Glu Pro
1625 1630 1635
Val Val Arg Leu Lys Asn Leu Thr Ser His Thr Lys Tyr Leu Val
1640 1645 1650
Ser Ile Ser Ala Phe Asn Ala Ala Gly Asp Gly Pro Lys Ser Asp
1655 1660 1665
Pro Gln Gln Gly Arg Thr His Gln Ala Ala Pro Gly Ala Pro Ser
1670 1675 1680
Phe Leu Ala Phe Ser Glu Ile Thr Ser Thr Thr Leu Asn Val Ser
1685 1690 1695
Trp Gly Glu Pro Ala Ala Ala Asn Gly Ile Leu Gln Gly Tyr Arg
1700 1705 1710
Val Val Tyr Glu Pro Leu Ala Pro Val Gln Gly Val Ser Lys Val
1715 1720 1725
Val Thr Val Glu Val Arg Gly Asn Trp Gln Arg Trp Leu Lys Val
1730 1735 1740
Arg Asp Leu Thr Lys Gly Val Thr Tyr Phe Phe Arg Val Gln Ala
1745 1750 1755
Arg Thr Ile Thr Tyr Gly Pro Glu Leu Gln Ala Asn Ile Thr Ala
1760 1765 1770
Gly Pro Ala Glu Gly Ser Pro Gly Ser Pro Arg Asp Val Leu Val
1775 1780 1785
Thr Lys Ser Ala Ser Glu Leu Thr Leu Gln Trp Thr Glu Gly His
1790 1795 1800
Ser Gly Asp Thr Pro Thr Thr Gly Tyr Val Ile Glu Ala Arg Pro
1805 1810 1815
Ser Asp Glu Gly Leu Trp Asp Met Phe Val Lys Asp Ile Pro Arg
1820 1825 1830
Ser Ala Thr Ser Tyr Thr Leu Ser Leu Asp Lys Leu Arg Gln Gly
1835 1840 1845
Val Thr Tyr Glu Phe Arg Val Val Ala Val Asn Glu Ala Gly Tyr

```
                1850                1855                1860
Gly Glu Pro Ser Asn Pro Ser Thr Ala Val Ser Ala Gln Val Glu
                1865                1870                1875
Ala Pro Phe Tyr Glu Glu Trp Trp Phe Leu Leu Val Met Ala Leu
                1880                1885                1890
Ser Ser Leu Ile Val Ile Leu Leu Val Val Phe Ala Leu Val Leu
                1895                1900                1905
His Gly Gln Asn Lys Lys Tyr Lys Asn Cys Ser Thr Gly Lys Gly
                1910                1915                1920
Ile Ser Thr Met Glu Glu Ser Val Thr Leu Asp Asn Gly Gly Phe
                1925                1930                1935
Ala Ala Leu Glu Leu Ser Ser Arg His Leu Asn Val Lys Ser Thr
                1940                1945                1950
Phe Ser Lys Lys Asn Gly Thr Arg Ser Pro Pro Arg Pro Ser Pro
                1955                1960                1965
Gly Gly Leu His Tyr Ser Asp Glu Asp Ile Cys Asn Lys Tyr Asn
                1970                1975                1980
Gly Ala Val Leu Thr Glu Ser Val Ser Leu Lys Glu Lys Ser Ala
                1985                1990                1995
Asp Ala Ser Glu Ser Glu Ala Thr Asp Ser Asp Tyr Glu Asp Ala
                2000                2005                2010
Leu Pro Lys His Ser Phe Val Asn His Tyr Met Ser Asp Pro Thr
                2015                2020                2025
Tyr Tyr Asn Ser Trp Lys Arg Arg Ala Gln Gly Arg Ala Pro Ala
                2030                2035                2040
Pro His Ser Val Ala Ile Leu Leu Thr Ser Asn Pro Ser Ala Tyr
                2045                2050                2055
Leu Ser Val Ala Pro Arg Gly Ser Ala Ser Trp
                2060                2065

<210> 6
<211> 333
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473826CD1

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

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

<210> 7
<211> 2700
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6471113CB1

<400> 7

```
atggctgtgc agcgcgccgc gtctccgcgc cgcccgcccg ccccgctctg gccccggctc 60
ctgctgccgc tgctgttgct gctgctgccc gcgccgagcg agggtcttgg ccactctgct 120
gaactggcat ttgctgtgga gccagtgat gatgttgccg tccccgggca gcctatagtg 180
ctggactgca gggtggaggg gaccctcca gtgcgaatca cctggaggaa gaatggggta 240
gagatgccag agagtaccca ctccaccttg ctggccaatg ggtccttgat gatccgtcac 300
ttcaggctgg agccgggagg cagccctcg gatgaaggtg actatgagtg tgtggcccag 360
aatcgctttg ggctggtgct cagccggaag gctcgcatcc aagctgcaac catgtcggac 420
ttccacgtgc atccccaggc caccgtgggt gaggagggtg gtgtggccgg cttccagtgc 480
caaatccatg ggcttcccaa accctgatc acttgggaga agaacagagt ccaattgac 540
acggacaatg agaggtacac attgctgccc aaggggctcc tgcagatcac aggacttcga 600
gctgaggacg gtggcatctt ccactgtgtg gcctcaaaca tcgccagtat ccggatcagc 660
cacgggccca ggctcactgt gtcaggctcg ggctctgggg cctacaggga gccagccatc 720
ctcgtggggc ctgaggaacc taccctgaca gtgcaccaga cgcggtgct tgagtgtgtc 780
gccacgggca acccgccgcc cattgtgtcc tggagccgcc tggatggtcg ccctatcggg 840
gtggagggca tccaggtgct gggcacagga aacctcatca tctcagacgt gacggtccag 900
```

cactctggcg tctacgtctg tgcagccaac agacctggca cccgggtgag gagaacggca 960
cagggccggc tggtggtgca agccccagct gagtttgtgc agcatcccca gtccatctcc 1020
aggccagctg ggaccacagc catgttcacc tgccaagccc agggtgagcc accgcctcat 1080
gtcacgtggc tgaaaaatgg acaggtgctg gggccaggag gccacgtcag gctcaagaat 1140
aacaacagca cactgaccat ttctggaatc ggtcctgagg atgcagccat ttatcagtgt 1200
gtggccgaga acagtgcggg ctcatcacag gccagtgcca ggctgaccgt actgtgggct 1260
gaggggctcc ccggccctcc ccgcaatgtg cgggcagtct ctgtgtcttc cactgacgtg 1320
cgtgtgtcct ggagtgagcc gctggccaac accaaggaga tcatcggcta cgtcctgcac 1380
atcaggaagg ctgctgaccc accggagctg cagtatcagg aggcagtcag caagagcacc 1440
tttcagcacc tggtcagcga cctggagccc tccacagcct acagtttcta catcaaggcc 1500
tacacaccaa gggggccag ctcagcctct gtgcccaccc tagctagcac cctgggtgaa 1560
gccctgccc cacccccact gtcagtgcga gtcctgggca gctcctcctt gcagctgctg 1620
tgggagcctt ggccccggct ggcccagcac gagggcggct tcaagctgtt ttacgcgcca 1680
gcaagcaaga cctccttcac cggccccatc ctgctgcctg gacccgtctc ctcctacaac 1740
ctcagccagc tcgaccccac tgcagtgtat gaggtgaagc tgctgcctca caaccagcat 1800
ggagatggca atgccacagt ccgctttgtg tctttgaggg gagcatctga gaggacaggc 1860
atcgtcatcg gcatccacat cgggctcact tgcatcatct tctgtgtcct cttcctcatg 1920
ttcggccaaa ggggcaggct cctcctgtgt aaagatgtgg aaaaccagct gtcccctcca 1980
cagggtcccc ggagccagag ggaccctggc attctggccc taaatggggc gagacgggga 2040
cagcggggcc agctgggccg agacgaraaa cgtgtggata tgaaggagct ggagcagctg 2100
ttccccccgg ccagccgcagc agggcagccg gaccccagac ccacagatcc tgcagccccc 2160
gctccgtgtg aggagaccca gctctcctg ctgccacttc aggggtgcgg cctgatggag 2220
gggagacgca cggaggcgaa gaccacagag gccacggctc cctgcgccgg cctggcggct 2280
gccccaccac ccccagatgg aggccctggc ctcctcagtg aaggccaggc ttccaggcct 2340
gcagccgccc gggttaccca gccagctcac tcggaacagt agccagtgtc tggcaggctc 2400
cagagggtgg acggagcggg gcccattctc aggtcaaaag caagatttct actgtcatgt 2460
gggatttgga tggtcctggg ggctccccag catttctatc ctgactgcct cttgggttgt 2520
caaaacccaa ggcagcttg acagggaccc cccggcccta acacccatca ggagttggag 2580
cagttcctgc aggagcctgt tcctccctg ggctgacgcc ccttgcccc tgcctgggta 2640
cccacatgac ttggaactga actaacattt ttctttaaaa agcaaaaact aaaaaaaaaa 2700

<210> 8
<211> 2321
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7701502CB1

<400> 8
gtataagtat atcccatgca gcatttcgga catacttata ccaaaaaaca taaaataaaa 60
tacttatat tgtctgtgaa agctttaacc aggcgtcctg cattttacct gacaagacca 120
gccccctcca gtaacggtcc cttcccctgt acccaccacc tccaccacca gtggcccaga 180
taagggtagc tcactgtgga cttctcactt ctgcttttcc gatgacccag ggcccagagg 240
gagagcccc acccccactac cccaccccca gccaggcccc caggatagac acaagcctgc 300
cctctcctcc tacacaggga gctctggaag gaccagctca acaagcccaa ggccaggctg 360
ggccccagg agaccagag gacaactggg caaggtgggc cggagagtgt gggggaaggc 420
aaaggagttc tgtgagctca gcgtctgaag gtcatttcat gcatcaggcc ccagggctca 480
gcttccgcct tcggcttccc cttctgccaa gagccctgag ccacttacag cacgaccaga 540
gaacaggcct gtctcaggca ggccctgcgc ctcctatgcg gagatgctac tgccactgct 600
gctgtcctcg ctgctgggcg ggtcccaggc tatggatggg agattctgga tacgagtgca 660
ggagtcagtg atggtgccgg agggcctgtg catctctgtg ccctgctctt tctcctaccc 720
ccgacaggac tggacagggt ctaccccagc ttatggctac tggttcaaag cagtgactga 780
gacaaccaag ggtgctcctg tggccacaaa ccaccagagt cgagaggtgg aaatgagcac 840
ccgggccga ttccagctca ctggggatcc cgccaagggg aactgctcct tggtgatcag 900

agacgcgcag atgcaggatg agtcacagtc cttctttcgg gtggagagag gaagctatgt 960
gagatataat ttcatgaacg atgggttctt tctaaaagta acagccctga ctcagaagcc 1020
tgatgtctac atccccgaga ccctggagcc cgggcagccg gtgacggtca tctgtgtgtt 1080
taactgggcc tttgaggaat gtccacccct ttctttctcc tggacggggg ctgccctctc 1140
ctcccaagga accaaaccaa cgacctccca cttctcagtg ctcagcttca cgcccagacc 1200
ccaggaccac aacaccgacc tcacctgcca tgtggactte tccagaaagg gtgtgagcgt 1260
acagaggacc gtccgactcc gtgtggccta tgccccaga gaccttgtta tcagcatttc 1320
acgtgacaac acgccagccc tggagcccca gcccaggga aatgtcccat accgtggagc 1380
ccaaaaggc cagttcctgc ggctcctctg tgctgctgac agccagcccc ctgccacact 1440
gagctgggtc ctgcagaaca gagtcctctc ctcgtcccat ccctgggccc ctagaccccct 1500
gggcctggag ctgcccgggg tgaaggctgg ggattcaggg cgctacacct gccgagcgga 1560
gaacaggctt ggctcccagc agcgagccct ggacctctct gtgcagtatc ctccagagaa 1620
cctgagagtg atggtttccc aagcaaacag gacagtcctg gaaaacctttg ggaacggcac 1680
gtctctccca gtactggagg gccaaagcct gtgcctggtc tgtgtcaaac acagcagccc 1740
cccagccagg ctgagctgga cccagaggag acaggttctg agcccctccc agccctcaga 1800
ccccgggttc ctggagctgc ctcgggttca agtggagcac gaaggagagt tcacctgcca 1860
cgctcggcac ccactgggct cccagcacgt ctctctcagc ctctccgtgc actataagaa 1920
gggactcatc tcaacgggat tctccaacgg agcgtttctg ggaatcggca tcacggctct 1980
tcttttcctc tgcctggccc tgatcatgta ggttaagagg aggcgtgggg gggtctgggg 2040
cctggaccca ggaagagggg aggtgttgca ggccgaaaga gtgaaggtcg tgatcaacgc 2100
agtatacctc tggaggttac atgagtaaac agcaaactgc tctcataaat gcagaatgtt 2160
gtccaactga caaactgcgt ctgcttccca gagggaatgc tgagggcagt cacgccccaa 2220
gcgaagtgtt tcttgtaatt aggcacagct gatgcttgtt agtaataata tgaacctgtg 2280
atcaattaaa cagctgacca attgttaaaa aaaaaaaaaa a 2321

<210> 9
<211> 2112
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5727379CB1

<400> 9
atgcttcaca cggccatatc atgctggcag ccattcctgg gtctggctgt ggtgttaatc 60
ttcatgggat ccaccattgg ctgccccgct cgctgtgagt gctctgccca gaacaaatct 120
gttagctgtc acagaaggcg attgatcgcc atcccagagg gcattcccat cgaaaccaaa 180
atcttggacc tcagtaaata caggctaaaa agcgtcaacc ctgaagaatt catatcatat 240
cctctgctgg aagagataga cttgagtgac aacatcattg ccaatgtgga accaggagca 300
ttcaacaatc tctttaacct gcttccctc cgcctaaaag gcaatcgtct aaagctggtc 360
cctttgggag tattcacggg gctgttccaat ctcactaagc ttgacattag tgagaatgag 420
attgttattt tactagacta catgttccaa gatctacata acctgaagtc tctagaagtg 480
ggggacaatg atttggttta tatatcacac agggcattca gtgggcttct tagcttggag 540
cagctcacct tggagaaatg caacttaaca gcagtaccaa cagaagcctt ctcccacctc 600
agcagcctca tcagcctgca tctgaagcat ctcaatatca acaatatgcc tgtgtatgcc 660
tttaaaagat tgttccacct gaaacaccta gagattgact attggcttt actggatatg 720
atgcctgcca atagcctcta cgttctcaac ctcacatccc tttcagtcac caacaccaat 780
ctgttactg taccttttct tgccttaaa cacctggagc ttcatatagt gggggccgag 840
cttgcacca ttgagcctca ctccttccaa gggctccgct tcctccgcgt gctcaatgtg 900
tctcagaacc tgctggaaac tttggaagag aatgtcttct cctccctag ggctctggag 960
gtcttgagca ttaacaacaa ccctctggcc tgtgactgcc gccttctctg gatcttgcag 1020
cgacagccca ccctgcagtt tggtggccag caaccttatgt gtgctggccc agcaccatc 1080
cgtgagaggt ctttcaagga tttccatagc actgccctt ctttttactt tacctgcaaa 1140
aaacccaaaa tccgtgaaaa gaagttgcag catctgctag tagatgaagg gcagacagtc 1200
cagctagaat gcagtgcaga tggagacccg cagcctgtga tttcctgggt gacacccgga 1260

aggcgtttca tcaccaccaa gtccaatgga agagccaccg tgttgggtga tggcaccttg 1320
gaaatccgct ttgcccagga tcaagacagc gggatgtatg tttgcatgcc tagcaatgct 1380
gctgggaatg atacttcac agcctcctta actgtgaaag gattcgcttc agatcgtttt 1440
ctttatgcga acaggaccca tatgtaaatg accgactcca atgaccaccat ttccaatggc 1500
accaatgcca atacttttttc cctggacctt aaaacaatac tggtgtctac agctatgggc 1560
tgcttcacat tcctgggagt ggttttattt tgttttcttc tccttttgt gtggagccga 1620
gggaaaggca agcacaaaaa cagcattgac cttgagtatg tgcccagaaa aaacaatggt 1680
gctgttgtgg aaggggaggt agctggaccc aggaggttca acatgaaaat gatttgaagg 1740
cccacccctc acattactgt ctctttgtca atgtgggtaa tcagtaagac agtatggcac 1800
agtaaattac tagattaaga ggcagccatg tgcagctgcc cctgtatcaa aagcagggtc 1860
tatggaagaa ggaggacttc caatggagac tctccatcga aaggcaggca ggcaggcatg 1920
tgtcagagcc cttcacacag tgggatacta agtgtttgcg ttgcaaatat tggcgttctg 1980
gggatcttag taatgaacct gaatatttgg ctcacactca cggacaatta ttcagcattt 2040
tctaccactg caaaaaaaaa aaaaaaaaat aaaaaagaac tacctacagt gtaggatttn 2100
catattaaaa ag 2112

<210> 10
<211> 1419
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5831801CB1

<400> 10
gtacgagttg gtcgagctcc atcactgata cggacgccag tgtgctggaa agcttacatg 60
taccctcctc ctcaggaaga aggagtgtta ggtggagaca gggatagttg ggacacaaag 120
tccagttaag tgtgtgtccc agctcagtgc tctgagccca gatcctcatc tcactggtga 180
gtggaggctca gcacagacaa gcaaccaact gctgggctgc cggtgccccc catgttggaa 240
cctgagttgg agattatctc ctaagcagat acctgtcttc caaactgggg atgtaggact 300
tggaaactag aaaatgccag gtctgaggga gaggaaagaa caagtccagc aatacacaga 360
gctctgtgta ttcagaggga agttggcagg gttgtgttcg ggcagagaaa ctccgagtgg 420
tacaaagggg acgtgcccag agtggagaaa tcatgctaat tgtctgcact agagctggag 480
aacgccaccc aaaatgaaga gagaaagggg agccctgtcc agagcctcca gggccctgcg 540
cattgctcct tttgtctacc ttctttctgat ccagacagac cccctggagg gggtgaacat 600
caccagcccc gtgcgcctga tccatgggac cgtgggaaag tcggctctgc tttctgtgca 660
gtacagcagt accagcagcg acaggcctgt agtgaagtgg cagctgaagc gggacaagcc 720
agtgaccgtg gtgcagtcca ttggcacaga ggtcatcggc accctgcggc ctgactatcg 780
agaccgtatc cgactctttg aaaatggctc cctgcttctc agcgacctgc agctggccga 840
tgagggcacc tatgaggtgg agatctccat caccgacgac accttcactg gggagaagac 900
catcaacctt actgtagatg tgccatttc gaggccacag gtgttggtgg cttcaaccac 960
tgtgctggag ctcagcgagg ccttcacctt gaactgctca catgagaatg gcaccaagcc 1020
cagctacacc tggctgaagg atggcaagcc cctcctcaat gactcgagaa tgctcctgtc 1080
cccgaccaa aaggtgctca ccatcaccag cgtgctcatg gaggatgacg acctgtacag 1140
ctgcatggtg gagaacccca tcagccaggg ccgcagcctg cctgtcaaga tcaccgtata 1200
cagaagaagc tcactttaca tcatcttgtc tacaggaggc atcttcctcc ttgtgacctt 1260
ggtgacagtc tgtgcctgct ggaaaccctc caaaaggaaa cagcagaagc tagaaaagca 1320
aaactccctg gaatacatgg atcagaatga tgaccgcctg aaaccagccc caaaggacca 1380
cagccctag gctggacaac tccttcagtt tgcccagcc 1419

<210> 11
<211> 6374
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7195090CB1

<400> 11
atggatgaag ctggaaacca tcattctcag caaactaaca caaggacaaa aaaccaaaca 60
ccgcatgttc tcactcatag gtacattatt ccatctttgc agaagctcga tgctgggttt 120
taccgctgcg tggtgcgaaa cagaatggga gcactcctgc aaagaaaatc agaagttcaa 180
gtcgcatata tgggaagttt catggatacg gaccagagga aaacagtttc tcaaggacgt 240
gcagcgattc taaacctgct gcccatcacc agctacccca gacctcaagt gacttggttt 300
agagaaggac acaagattat tccaagcaac agaatagcca tcacattgga gaatcagctg 360
gtgatcctcg ccaccacaac cagtgatgcc ggggcatact acgtgcaggc cgtgaatgag 420
aaaaatggag aaaacaagac aagcccattc attcatttga gcatagcaag agatgttggc 480
acacctgaaa ccatggcccc aaccattgtg gttccccggg gcaacagaag tgtggtggct 540
ggatccagtg agaacacctt ggaatgtata gccagtgcca ggcctgtgga ggacctgagt 600
gtgacctgga agaaggaatgg agtgagaatc accagtggcc tccacagctt tggaagacgc 660
ctcatcatca gcaacccgac gtccgcggac accgggccat acgtctgcga ggcggcgctg 720
ccggggagcg cttttgaacc ggccagggcg acggcctttc ttttcatcat agagccacca 780
tattttactg ctgagcccga gagtcggatt tcagatgaag tagaagaaac tgtggacatc 840
ggatgtcaag ccatgggggt cccccttccc accctccagt ggtacaagga tgccatctcc 900
atcagcaggc tccagaatcc tcgatacaaa gtgctcgcca gcggaggcct gcgcatccag 960
aagctgcgtc cagaggactc cggaatcttc cagtgcttcg ccagcaatga aggaggggag 1020
atccagacct acacctacct ggatgtaacc aatatcgctc cagtgtttac ccagcgggca 1080
gtggacacca cagttactga cgtgatgacc gccattcaaa ggtgtgaggt gtccgggact 1140
cccaaacccg ccatcacctg gaaaagagaa aaccacattc tggccagtgg ctctgtccgg 1200
attcctaggt tcatgcttct tgaatcgggg ggtctacaga tcgcgccgct cttcatccag 1260
gatgccggca actacacctg ctatgcggcc aacacagagg gctcctgaa tgcatcggcc 1320
aagctcactg tgtggaatcg gacgtccatc gtccacccta ctgaggacca cgtggtgatt 1380
aagggggacc cggccacgct gcactgtggt gccacacatg acccccgggt ttcactccgc 1440
tacgtttgga agaaggacaa cgtggccctg atccatcgga gcacgtctag gatcgtggtg 1500
gagaaggacg ggtccttct catcagccag acgtggtcag gcgacatcgg tgactacagc 1560
tgcgagattg tttctgaagg agggaatgac tccaggatgg cccggctgga agtgattgaa 1620
ctgcctcatt cacctcagaa cctcctggtc agccctaatt attccacag ccacgccgtg 1680
gtgctctctt gggtccggcc ctttgatgga aacagtccta ttctttatta catcgtggag 1740
ctctctgaaa acaactccgc atggaaggtg catctgtcaa acgttggccc tgagatgaca 1800
ggcgtcaccg tgagtggcct gactccggct cgtacctatc aattccgggt gtgcgcggtg 1860
aatgaagtgg gcagggccca gtacagcgcc gagacaagca ggttgatgct acctgaagaa 1920
ccacccagtg ctccccgaa aaatatagtg gccagtgggc ggactaatca gtccattatg 1980
gtccagtggc agccacccc agaaacagag cacaacgggg tgttgcgtgg atacatcctc 2040
aggtaccgcc tggctggcct tcccggagag taccagcagc ggaacatcac cagcccggag 2100
gtgaactact gcctggtgac agacctgatc atcttggacac agtatgagat acaggtggcg 2160
gcgtacaacg gggccggtct gggcgtcttc agcagggcag tgaccgagta cacattgcag 2220
ggagtgccca ccgcgcccc gcagaacgtg cagacggaag ccgtgaactc caccaccatt 2280
cagttcctgt ggaacctcac gcctcagcag tttatcaatg gcatcaacca gggatacaag 2340
cttctggcat ggccggcaga tgccccgag gctgtcactg tggtcactat tgccccagat 2400
ttccacggag tccaccatgg acacataacg aacctgaaga agtttaccgc ctacttcact 2460
tccgttctgt gcttcaccac ccctgcggac gggcctccca gcacacctca gctggtctgg 2520
actcaggaag acaaaaccagg agctgtggga catctgagtt tcacagagat cttggacaca 2580
tctctcaagg tcagctggca ggagccctg gagaaaaatg gcatcattac tggctatcag 2640
atctcttggg aagtgtacgg caggaacaac tctcgtctca cgccacaccct gaacagcacg 2700
acgcacgagt acaagatcca aggcctctca tctctcacca cctacaccat cgacgtggcc 2760
gctgtgactg ccgtgggcac tggcctggtg atttcatcca ccatttcttc tgtagtgccc 2820
ccagacctcc ctggtgccca atccaacctg gtcatttcca acatcagccc tcgctccgcc 2880
accctteagt tccggccagg ctatgacggg aaaacgtcca tctccaggtg gattgttgag 2940
gggcaggtgg gagctatcgg cgacgagyag gagtgggtca cctctatga agaggagaat 3000
gagcctgatg cccagatgct ggagatccca aacctcaaac cctacattca ctacagattt 3060

```
cgaatgaagc aagtgaacat tgttgggccg agcccctaca gtccgtcttc ccgggtcatc 3120
cagaccctgc agccccacc cgacgtggct ccaacgagcg tcacggtccg tactgccagt 3180
gagaccagcc tgcggcttcg ctgggtgccc ctgccggatt ctcagtacaa cgggaacccc 3240
gagtccgtgg gctacaggat taagtactgg cgctcagacc tccagtcctc agccagtggcc 3300
cnagtcgtca gtgaccggct ggagagagaa ttcaccatcg aggagctgga ggagtggatg 3360
gaatacgagc tgcagatgca ggccttcaac gccgtcgggg ctgggccgtg gagcgaggtg 3420
gtgcgggccc ggacgcggga gtcagttcct tcagccgccc ctgagaacgt gtcagccgag 3480
gctgtcagct cgacccagat tttactgaca tggacatccg tgccggaaca ggaccagaat 3540
gggctcatac tgggctacaa gatcctgttc cgggccaaag acctggatcc cgagcccagg 3600
agccacatcg tgcgagggaa ccacacgcag tcggccctgc tggcaggcct gcgcaagttc 3660
gtgctctacg agctccaggt gctggcgttc accgccatcg ggaacggggt ccccagcacg 3720
cccctcatcc tgghgcgcac caaagacgat gcccccaggcc caccagtgag gctcgtgttc 3780
cccgaagtga gactcacctc cgtgcggata gtgtggcaac ctccggagga gcccaacggc 3840
atcatcctgg ggtcccagat tgcctaccgc ctggccagca gcagcccccа cacctcacc 3900
accgtggagg tcggcgccac agtgaggcag ttcacagcca ccgacctggc cccggagtcc 3960
gcatacatct tcaggctgtc cgccaagacg aggcaggct ggggggagcc actggaggcc 4020
accgtcatca ccacgagaa gagagagcgg ccggcacccc ccagagagct cctggtgccc 4080
caggcagaag tgaccgcacg cagcctccgg ctccagtggg tcccgggcag cgacgggcc 4140
tccccatcc ggtacttcac catgcaggtg cgagagctgc ctcgggtga gtggcagacc 4200
tactcctcgt ccatcagcca tgaggcgaca gcatgcgtcg ttgacagact gaggcccttc 4260
acctcctaca agctgcgcct gaaagccacc aacgacattg gggacagtga cttcagttca 4320
gagacagagg cggtgaccac gctgcaggat gttccaggag agcccccggg atctgtctca 4380
gcgacgccac acaccacgtc ctctgtcctg atacagtggc agcctccgag ggacgaaagc 4440
ctgaatggcc ttcttcaggg atacaggatc tactacaggg agctggagta tgaagccggg 4500
tcaggcactg aggccaagac gctcaaaaac cctatagctt tacatgctga gctcacagcc 4560
caaagcagct tcaagacggt gaacagcagc tccacatcga cgatgtgtga actaacacat 4620
ttaaagaagt acggcgcgta tgaagtaata atgaccgcct ataccatcat cggcgagagc 4680
ccagccagcg cgccgtgga ggtctttgtc ggcgaggctg ccacggccat ggcccgcgag 4740
aacgtgcagg tgaccccact cacggccagc cagctggagg tcatgtggga cccaccaccc 4800
ccggagagcc agaatgggaa catccaaggc tacaagattt actactgggа ggcagacagc 4860
cagaacgaaa cggagaaaat gaaggtcctc ttcctccccg agcccgtggt gaggctgaag 4920
aacctgacca gccatgccaa gtacctggtc agcatatcag ccttcaacgc cgccggagat 4980
ggacctaaga gtgacccccа gcagggcgcc acccaccagg ccgcccctgg ggcccccagc 5040
tttctggcgt tctcagaaat aacctcnacc acgctcaacg tgtcctgggg cgagcctgcg 5100
gcggccaacg gcatcctgca gggctatcgg gtggtgtacg agcccttggc cctgtacaa 5160
ggggtgagca aggtgntgac cgtggaagtg agagggaact ggcagcgctg gctgaaggtg 5220
cgggacctca ccaagggagt gacctatttc tccgtgtcc aagcgcggac catcacctac 5280
gggcccgagc tccnagccaa tatcacagcc gggccagccg agggatcccc gggctcgcct 5340
agagatgtcc tggtcaccaa gtccgcttct gaactgacgc tgcagtggac tgagggacac 5400
tctggcgaca cacctaccac gggctatgtg atcgaggccc ggcctcaga tgaaggctta 5460
tgggacatgt ttgtgaagga catcccgcgg agcgccacat cctacaccct cagcctggat 5520
aagctccggc aaggagtgac ttacgagttc cgggtggtgg ctgtgaatga ggcgggctac 5580
ggggagccca gcaacccctc cacggctgtg tcagctcaag tggaagcccc attctacgag 5640
gagtggtggt tcttcttggt gatggctctg tccagcctga tcgtcatcct gctggtggtg 5700
ttcgccctcg tcttgcacgg gcagaataag aagtataaga actgcagcac aggaaagggg 5760
atctcaacca tggaggagtc tgtgaccctg gacaacggag gatttgctgc cctggagctc 5820
agcagccgcc acctcaatgt caagagcacc ttcttcaaga agaacgggac caggtcccca 5880
cccggcctа gcccggcgg cctgcactac tcagacgagg acatctgcaa cnagtacaac 5940
ggcgccgtgc tgaccgagcg cgtgagcctc aagyagaagt cggcagatgc atcagaatct 6000
gaggccacgg actttgacta cgaggacgcg ctgcccaagc actccttcgt gaacgacbac 6060
atgagcgacc ccnncctacta caacncatgg aagcgcaggg cccagggccg cgcacctgcg 6120
ccgcacagtg tagctatcct cctgacgagc aacccgtctg cgtacctaag tgtggctccc 6180
cgtgggtcag cgtcctggta gcatggatcc agtctgaaag gtgaggacaa cgtggaaact 6240
catgagctga gcctgcccgc tgggacacgt ctccttcccg cgtcaccttc tggtttaggg 6300
agccgtcagg tcnctaaacg ttcnctacaa cttttttctga aattgtgcag aaaaacagat 6360
ctcattaaga aaaa 6374
```

<210> 12
<211> 1350
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473826CB1

<400> 12
atgacaagcg tgaccataaa aggctctgga tttgccgttt cttctgcagg tgtaaaagtc 60
cttatgggtc atttcccatg taaagttcta tcagtgaatt atacggccat tgaatgtgaa 120
acatcccctg ctgcccaaca gcttgtggat gtagatcttc taatacatgg agtgcctgcc 180
cagtgccagg gaaactgcac cttttcatac ttagaaagca tcactcctta cataacagga 240
gtcttcccaa actctgtcat aggatctgta aaagttctta ttgaaggaga aggtttgggg 300
actgttttgg aggacatttt tggacatcat tctgttagtg ttgtggtggg aagtaaaggc 360
ttggctctgg gaaacctgac tgtcagcagc ccccagtag catctctatc acaaacttct 420
ggaagcattg gtggtggaac tacactggtg atcacaggaa atggcttcta tccaggcaac 480
actacagtca ctattgggga tgaaccttgt caaatttatt ccatcaaccc caatgaagtc 540
tactgccgca ctcccgctgg gaccactgga atggtcgatg ttaaaatctt tgttaataca 600
attgcttatc caccttgct ttttacatat gcctggagg atactccatt tctcagagga 660
attatcccaa gcagaggtcc accaggaact gcaattgaga tcactggatc caactttggc 720
tttgagatct tggaaatctc cgtgattgata aataacattc agtgtaatgt aaccatggcc 780
aatgatagtg tgttgcagtg catggtggga gatcatgctg ggggcacatt tcttgttatg 840
atgcatcaaa agacaaaagg ctcagccacg tccacagttg tatttgagaa cccgcttaat 900
attcaaaata ttaatccaag ccagggtagt cataagtatg caagaacctg cagatatttt 960
gtttcagtaa acacttattt agaaaacttа tttagaaaat aatatgttac tcaaacacat 1020
aaaatttcta gtgttttttt tgtttgtttg gttttttttt tttgagacag agtctcgctc 1080
ggtcgcccag gctggagtgc agtggcgcga tctttgctca ttgcaagctc tgcctcccag 1140
gttcacgcca ttctcctgcc tcagcctccc gagtagctgg gactacaggc gcccgccacc 1200
atgcctggct aattttttt ggtatttta gtagagacgg ggtttcacca tgaaatttta 1260
tagcacaata caaatacttg gctttatgta gcaaataaat aagcttgtaa gtgtataatt 1320
atttatttca aaatactaaa ctttaaaata 1350

【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
23 May 2002 (23.05.2002)

PCT

(10) International Publication Number
WO 2002/040671 A3

(51) International Patent Classification⁷: C12N 15/12, C07K 14/705, 16/28, A01K 67/027, A61K 38/16, C12N 5/10, G01N 33/50, 33/53

(21) International Application Number:
PCT/US2001/044974

(22) International Filing Date:
15 November 2001 (15.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/249,645 16 November 2000 (16.11.2000) US

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

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): BAUGHN, Mariah, R. [US/US]; 1424 Sandago Road, San Leandero, CA 94577 (US). LU, Dyung, Aina, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). YUE, Henry [US/US]; 826 Luis Avenue, Sunnyvale, CA 94087 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). THANGAVELU, Kavitha [IN/US]; 1950 Montecito Avenue, #23, Mountain View, CA 94043 (US). RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). LU, Yan [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). LO, Terrence, P. [CA/US]; 1451 Beach Park Boulevard, #115, Foster City, CA 94404 (US). GURURAJAN, Rajagopal [IN/US]; 5591 Dent Avenue, San Jose, CA 95118 (US). GANDHI, Ameena, R. [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). ARVIZU, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). YAO, Monique, G. [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (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, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, 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:
31 December 2003

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 2002/040671 A3

(54) Title: IMMUNOGLOBULIN SUPERFAMILY PROTEINS

(57) Abstract: The invention provides human immunoglobulin superfamily proteins (IGSFP) and polynucleotides which identify and encode IGSFP. 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 IGSFP.

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/44974

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C07K14/705 C07K16/28 A01K67/027 A61K38/16
C12N5/10 G01N33/50 G01N33/53

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
BIOSIS, EPO-Internal, EMBL, SEQUENCE SEARCH, WPI Data, PAJ, MEDLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SALBAUM J MICHAEL: "Punc, a novel mouse gene of the immunoglobulin superfamily, is expressed predominantly in the developing nervous system." MECHANISMS OF DEVELOPMENT, vol. 71, no. 1-2, February 1998 (1998-02), pages 201-204, XP002215120 ISSN: 0925-4773 figure 1 --- -/--	1-56,62

[X] Further documents are listed in the continuation of box C. [X] Patent family members are listed in annex.

* Special categories of cited documents :
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 September 2002	17.01.2003

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Schmitz, T

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

Inte... Application No
PCT/US 01/44974

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] 25 April 1998 (1998-04-25) SALBAUM, J.: "Mus musculus putative neuronal cell adhesion molecule (Punc) mRNA, complete cds" retrieved from EBI Database accession no. AF026465 XP002215121 abstract ---	1-56,62
P,X	WO 01 66747 A (VERNET CORINE A M ;MEZES PETER S (US); MISHRA VISHNU (US); RASTELL) 13 September 2001 (2001-09-13) SEQ ID NO:59 page 120; table 8B page 29 -page 30; table 2P ---	1-56,62
A	WO 00 29583 A (INCYTE PHARMA INC ;CORLEY NEIL C (US); YUE HENRY (US); BAUGHN MARI) 25 May 2000 (2000-05-25) the whole document ---	
A	SALBAUM J M: "GENOMIC STRUCTURE AND CHROMOSOMAL LOCALIZATION OF THE MOUSE GENE PUNC" MAMMALIAN GENOME, NEW YORK, NY, US, vol. 10, no. 2, February 1999 (1999-02), pages 107-111, XP000992754 ISSN: 0938-8990 the whole document -----	

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

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 01/44974

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

see FURTHER INFORMATION sheet PCT/ISA/210

2. [X] Claims Nos.: 13, 21, 22, 24, 25 (all partially)
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

see FURTHER INFORMATION sheet PCT/ISA/210

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

see additional sheet

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

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

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

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-55 (partially); 56, 62 (complete)

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (1)) (July 1998)

International Application No. PCT/US 01/44974

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claims 19, 22, 25 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Although claims 33, 35 are directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Continuation of Box I.2

Claims Nos.: 13, 21, 22, 24, 25 (all partially)

Present claims 21, 22, 24, 25 relate to an agonist / antagonist defined by reference to a desirable characteristic or property, namely having "agonist-activity" or "antagonist-activity".
The claims cover all agonists / antagonists having this characteristic or property, whereas the application provides support within the meaning of Article 6 PCT and disclosure within the meaning of Article 5 PCT for only a very limited number of such compounds. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT). An attempt is made to define the agonist / antagonist by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search over the whole of the claimed scope impossible. Consequently, the search has been carried out for those parts of the claims which appear to be clear, supported and disclosed, namely those parts relating to the antibody binding to the claimed polypeptide.

Present claim 13 relates to an extremely large number of possible nucleotide sequences. Support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT is to be found, however, for only a very small proportion of the sequences claimed. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Consequently, the search has been carried out for those parts of the claims which appear to be supported and disclosed, namely those parts relating to the full length of the sequence (SEQ ID NO:7).

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

receipt of the search report or during any Chapter II procedure.

International Application No. PCT/US 01/44974

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

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

1. Claims: 1-55 (partially), 56, 62 (completely)

An isolated polypeptide (SEQ ID NO:1) together with the coding DNA (SEQ ID NO:7).
Furthermore vectors, host cells, methods, diagnostic tests, arrays, compositions, antibodies, all referring to said nucleotide or amino acid sequence.

2. Claims: 1-55 (partially), 57, 63 (completely)

As subject 1, but referring to SEQ ID NO:2, 8.

3. Claims: 1-55 (partially), 58, 64 (completely)

As subject 1, but referring to SEQ ID NO:3, 9.

4. Claims: 1-55 (partially), 59, 65 (completely)

As subject 1, but referring to SEQ ID NO:4, 10.

5. Claims: 1-55 (partially), 60, 66 (completely)

As subject 1, but referring to SEQ ID NO:5, 11.

6. Claims: 1-55 (partially), 61, 67 (completely)

As subject 1, but referring to SEQ ID NO:6, 12.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int al Application No
PCT/US 01/44974

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 0166747 A	13-09-2001	AU	4729401 A	17-09-2001
		EP	1261712 A2	04-12-2002
		WO	0166747 A2	13-09-2001
		US	2002155115 A1	24-10-2002
		EP	1169445 A2	09-01-2002
		WO	0061754 A2	19-10-2000
		US	2002172999 A1	21-11-2002
		US	2002168716 A1	14-11-2002
WO 0029583 A	25-05-2000	AU	1631000 A	05-06-2000
		EP	1131428 A2	12-09-2001
		JP	2002530080 T	17-09-2002
		WO	0029583 A2	25-05-2000

Form PCT/ISA/210 (patent family annex) (July 1992)

フロントページの続き

(51)Int.Cl.[7]	FI		テーマコード（参考）
Ａ６１Ｐ 1/18	Ａ６１Ｐ 1/18		４Ｂ０６５
Ａ６１Ｐ 3/10	Ａ６１Ｐ 3/10		４Ｃ０８４
Ａ６１Ｐ 5/14	Ａ６１Ｐ 5/14		４Ｃ０８５
Ａ６１Ｐ 5/38	Ａ６１Ｐ 5/38		４Ｈ０４５
Ａ６１Ｐ 5/40	Ａ６１Ｐ 5/40		
Ａ６１Ｐ 7/00	Ａ６１Ｐ 7/00		
Ａ６１Ｐ 7/06	Ａ６１Ｐ 7/06		
Ａ６１Ｐ 7/10	Ａ６１Ｐ 7/10		
Ａ６１Ｐ 9/00	Ａ６１Ｐ 9/00		
Ａ６１Ｐ 9/04	Ａ６１Ｐ 9/04		
Ａ６１Ｐ 9/10	Ａ６１Ｐ 9/10		
Ａ６１Ｐ 11/00	Ａ６１Ｐ 11/00		
Ａ６１Ｐ 11/06	Ａ６１Ｐ 11/06		
Ａ６１Ｐ 13/12	Ａ６１Ｐ 13/12		
Ａ６１Ｐ 15/00	Ａ６１Ｐ 15/00		
Ａ６１Ｐ 17/00	Ａ６１Ｐ 17/00		
Ａ６１Ｐ 17/06	Ａ６１Ｐ 17/06		
Ａ６１Ｐ 19/02	Ａ６１Ｐ 19/02		
Ａ６１Ｐ 19/06	Ａ６１Ｐ 19/06		
Ａ６１Ｐ 19/10	Ａ６１Ｐ 19/10		
Ａ６１Ｐ 21/00	Ａ６１Ｐ 21/00		
Ａ６１Ｐ 21/04	Ａ６１Ｐ 21/04		
Ａ６１Ｐ 25/00	Ａ６１Ｐ 25/00		
Ａ６１Ｐ 25/02	Ａ６１Ｐ 25/02		
Ａ６１Ｐ 25/08	Ａ６１Ｐ 25/08		
Ａ６１Ｐ 25/14	Ａ６１Ｐ 25/14		
Ａ６１Ｐ 25/16	Ａ６１Ｐ 25/16		
Ａ６１Ｐ 25/18	Ａ６１Ｐ 25/18		
Ａ６１Ｐ 25/20	Ａ６１Ｐ 25/20		
Ａ６１Ｐ 25/28	Ａ６１Ｐ 25/28		
Ａ６１Ｐ 27/02	Ａ６１Ｐ 27/02		
Ａ６１Ｐ 27/06	Ａ６１Ｐ 27/06		
Ａ６１Ｐ 27/12	Ａ６１Ｐ 27/12		
Ａ６１Ｐ 27/16	Ａ６１Ｐ 27/16		
Ａ６１Ｐ 29/00	Ａ６１Ｐ 29/00	１０１	
Ａ６１Ｐ 31/04	Ａ６１Ｐ 31/04		
Ａ６１Ｐ 31/10	Ａ６１Ｐ 31/10		
Ａ６１Ｐ 31/12	Ａ６１Ｐ 31/12		
Ａ６１Ｐ 31/18	Ａ６１Ｐ 31/18		
Ａ６１Ｐ 33/00	Ａ６１Ｐ 33/00		
Ａ６１Ｐ 33/02	Ａ６１Ｐ 33/02		
Ａ６１Ｐ 35/00	Ａ６１Ｐ 35/00		
Ａ６１Ｐ 35/02	Ａ６１Ｐ 35/02		
Ａ６１Ｐ 37/02	Ａ６１Ｐ 37/02		
Ａ６１Ｐ 37/04	Ａ６１Ｐ 37/04		
Ａ６１Ｐ 37/06	Ａ６１Ｐ 37/06		
Ａ６１Ｐ 37/08	Ａ６１Ｐ 37/08		

Ｃ０７Ｋ	1/22	Ｃ０７Ｋ	1/22
Ｃ０７Ｋ	14/47	Ｃ０７Ｋ	14/47
Ｃ０７Ｋ	16/18	Ｃ０７Ｋ	16/18
Ｃ０７Ｋ	19/00	Ｃ０７Ｋ	19/00
Ｃ１２Ｍ	1/00	Ｃ１２Ｍ	1/00 Ａ
Ｃ１２Ｎ	1/15	Ｃ１２Ｎ	1/15
Ｃ１２Ｎ	1/19	Ｃ１２Ｎ	1/19
Ｃ１２Ｎ	1/21	Ｃ１２Ｎ	1/21
Ｃ１２Ｎ	5/10	Ｃ１２Ｐ	21/02 Ｃ
Ｃ１２Ｎ	15/02	Ｃ１２Ｐ	21/08
Ｃ１２Ｐ	21/02	Ｃ１２Ｑ	1/02
Ｃ１２Ｐ	21/08	Ｃ１２Ｑ	1/48 Ｚ
Ｃ１２Ｑ	1/02	Ｃ１２Ｑ	1/68 Ａ
Ｃ１２Ｑ	1/48	Ｇ０１Ｎ	33/15 Ｚ
Ｃ１２Ｑ	1/68	Ｇ０１Ｎ	33/50 Ｚ
Ｇ０１Ｎ	33/15	Ｇ０１Ｎ	33/53 Ｄ
Ｇ０１Ｎ	33/50	Ｇ０１Ｎ	33/53 Ｍ
Ｇ０１Ｎ	33/53	Ｇ０１Ｎ	37/00 １０２
Ｇ０１Ｎ	37/00	Ｃ１２Ｎ	15/00 Ｃ
		Ｃ１２Ｎ	15/00 Ｆ
		Ｃ１２Ｎ	5/00 Ａ

(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,PL,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZW

(72)発明者 ユエ、ヘンリー
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
(72)発明者 エリオット、ビッキー・エス
アメリカ合衆国カリフォルニア州９５１２１・サンノゼ・ポルトンプレイスウェイ ３７７０
(72)発明者 サンガベル、カビサ
アメリカ合衆国カリフォルニア州９４０４３・マウンテンビュー・＃２３・モンテシトアベニュー １９５０
(72)発明者 ランクマール、ジャヤラクシミ
アメリカ合衆国カリフォルニア州９４５５５・フレモント・メイバードサークル ３４３５９
(72)発明者 リュ、ヤン
アメリカ合衆国カリフォルニア州９４３０３・パロアルト・コリーナウェイ ３８８５
(72)発明者 ロー、テレンス・ピー
アメリカ合衆国カリフォルニア州９４４０４・フォスターシティ・＃１１５・ビーチパークブールバード １４５１
(72)発明者 ガルラジャン、ラジャゴパル
アメリカ合衆国カリフォルニア州９５１１８・サンノゼ・デントアベニュー ５５９１
(72)発明者 ガンディー、アミーナ・アール
アメリカ合衆国カリフォルニア州９４１１８・サンフランシスコ・フィフスアベニュー ７０５
(72)発明者 アービズ、チャンドラ
アメリカ合衆国カリフォルニア州９４０２５・メンロパーク・＃１・シャーウッドウェイ ４９０
(72)発明者 ヤオ、モニーク・ジー
アメリカ合衆国インディアナ州４６０３３・カーメル・ウッドゲートドライブ １１８９

Ｆターム(参考) 2G045 AA25 AA40 BB03 BB20 CB01 CB21 DA12 DA13 DA14 DA36
DA37 DA77 FB02 FB03
4B024 AA01 AA11 BA41 BA61 BA63 CA04 DA02 DA06 EA04 HA01
HA04 HA14
4B029 AA07 BB20 CC03 FA15
4B063 QA01 QA05 QA18 QQ42 QR07 QR32 QR56 QR82 QS25 QS34
QS36 QX02 QX07
4B064 AG20 AG27 CA02 CA10 CA19 CA20 CC24 CE12 CE13 DA01
DA13
4B065 AA26X AA90X AA93Y AB01 BA02 CA24 CA25 CA44 CA46
4C084 AA17 NA14 ZA012 ZA022 ZA052 ZA062 ZA152 ZA162 ZA182 ZA202
ZA222 ZA332 ZA342 ZA362 ZA402 ZA452 ZA512 ZA552 ZA592 ZA662
ZA682 ZA752 ZA812 ZA832 ZA892 ZA942 ZA962 ZA972 ZB052 ZB082
ZB092 ZB132 ZB152 ZB262 ZB272 ZB322 ZB332 ZB352 ZB382 ZB392
ZC042 ZC062 ZC082 ZC312 ZC352 ZC552
4C085 KA03 KA04 KA05 LL03 LL07 LL13 LL17 LL18
4H045 AA10 AA11 AA20 AA30 BA10 BA41 CA40 DA50 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	JP2004535759A	公开(公告)日	2004-12-02
申请号	JP2002543666	申请日	2001-11-15
[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	ボーグンマライアアール リュデュングアイナエム ユエヘンリー エリオットビッキーエス サンガベルカビサ ランクマールジャヤラクシミ リュヤン ローテレンスピー ガルラジャンラジャゴパル ガンディーアミーナアール アービズチャンドラ ヤオモニークジー		
发明人	ボーグン、マライア·アール リュ、デュング·アイナ·エム ユエ、ヘンリー エリオット、ビッキー·エス サンガベル、カビサ ランクマール、ジャヤラクシミ リュ、ヤン ロー、テレンス·ピー ガルラジャン、ラジャゴパル ガンディー、アミーナ·アール アービズ、チャンドラ ヤオ、モニーク·ジー		
IPC分类号	G01N33/50 A61K45/00 A61K49/00 A61P1/04 A61P1/16 A61P1/18 A61P3/10 A61P5/14 A61P5/38 A61P5/40 A61P7/00 A61P7/06 A61P7/10 A61P9/00 A61P9/04 A61P9/10 A61P11/00 A61P11/06 A61P13/12 A61P15/00 A61P17/00 A61P17/06 A61P19/02 A61P19/06 A61P19/10 A61P21/00 A61P21/04 A61P25/00 A61P25/02 A61P25/08 A61P25/14 A61P25/16 A61P25/18 A61P25/20 A61P25/28 A61P27/02 A61P27/06 A61P27/12 A61P27/16 A61P29/00 A61P31/04 A61P31/10 A61P31/12 A61P31/18 A61P33/00 A61P33/02 A61P35/00 A61P35/02 A61P37/02 A61P37/04 A61P37/06 A61P37/08 C07K1/22 C07K14/47 C07K14/705 C07K16/18 C07K19/00 C12M1/00 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N15/02 C12N15/09 C12N15/12 C12P21/02 C12P21/08 C12Q1/02 C12Q1/48 C12Q1/68 C12Q1/6883 G01N33/15 G01N33/53 G01N37/00		
CPC分类号	A01K2217/05 A61P1/04 A61P1/16 A61P1/18 A61P11/00 A61P11/06 A61P13/12 A61P15/00 A61P17/00 A61P17/06 A61P19/02 A61P19/06 A61P19/10 A61P21/00 A61P21/04 A61P25/00 A61P25/02 A61P25/08 A61P25/14 A61P25/16 A61P25/18 A61P25/20 A61P25/28 A61P27/02 A61P27/06 A61P27/12 A61P27/16 A61P29/00 C07K14/70503 C12Q1/6883 C12Q2600/158		
FI分类号	C12N15/00.A A61K45/00 A61K49/00.A A61P1/04 A61P1/16 A61P1/18 A61P3/10 A61P5/14 A61P5/38 A61P5/40 A61P7/00 A61P7/06 A61P7/10 A61P9/00 A61P9/04 A61P9/10 A61P11/00 A61P11/06 A61P13/12 A61P15/00 A61P17/00 A61P17/06 A61P19/02 A61P19/06 A61P19/10 A61P21/00 A61P21/04 A61P25/00 A61P25/02 A61P25/08 A61P25/14 A61P25/16 A61P25/18 A61P25/20 A61P25/28 A61P27/02 A61P27/06 A61P27/12 A61P27/16 A61P29/00.101 A61P31/04 A61P31/10 A61P31/12		

	A61P31/18 A61P33/00 A61P33/02 A61P35/00 A61P35/02 A61P37/02 A61P37/04 A61P37/06 A61P37/08 C07K1/22 C07K14/47 C07K16/18 C07K19/00 C12M1/00.A C12N1/15 C12N1/19 C12N1/21 C12P21/02.C C12P21/08 C12Q1/02 C12Q1/48.Z C12Q1/68.A G01N33/15.Z G01N33/50.Z G01N33/53.D G01N33/53.M G01N37/00.102 C12N15/00.C C12N15/00.F C12N5/00.A
F-TERM分类号	2G045/AA25 2G045/AA40 2G045/BB03 2G045/BB20 2G045/CB01 2G045/CB21 2G045/DA12 2G045/DA13 2G045/DA14 2G045/DA36 2G045/DA37 2G045/DA77 2G045/FB02 2G045/FB03 4B024/AA01 4B024/AA11 4B024/BA41 4B024/BA61 4B024/BA63 4B024/CA04 4B024/DA02 4B024/DA06 4B024/EA04 4B024/HA01 4B024/HA04 4B024/HA14 4B029/AA07 4B029/BB20 4B029/CC03 4B029/FA15 4B063/QA01 4B063/QA05 4B063/QA18 4B063/QQ42 4B063/QR07 4B063/QR32 4B063/QR56 4B063/QR82 4B063/QS25 4B063/QS34 4B063/QS36 4B063/QX02 4B063/QX07 4B064/AG20 4B064/AG27 4B064/CA02 4B064/CA10 4B064/CA19 4B064/CA20 4B064/CC24 4B064/CE12 4B064/CE13 4B064/DA01 4B064/DA13 4B065/AA26X 4B065/AA90X 4B065/AA93Y 4B065/AB01 4B065/BA02 4B065/CA24 4B065/CA25 4B065/CA44 4B065/CA46 4C084/AA17 4C084/NA14 4C084/ZA012 4C084/ZA022 4C084/ZA052 4C084/ZA062 4C084/ZA152 4C084/ZA162 4C084/ZA182 4C084/ZA202 4C084/ZA222 4C084/ZA332 4C084/ZA342 4C084/ZA362 4C084/ZA402 4C084/ZA452 4C084/ZA512 4C084/ZA552 4C084/ZA592 4C084/ZA662 4C084/ZA682 4C084/ZA752 4C084/ZA812 4C084/ZA832 4C084/ZA892 4C084/ZA942 4C084/ZA962 4C084/ZA972 4C084/ZB052 4C084/ZB082 4C084/ZB092 4C084/ZB132 4C084/ZB152 4C084/ZB262 4C084/ZB272 4C084/ZB322 4C084/ZB332 4C084/ZB352 4C084/ZB382 4C084/ZB392 4C084/ZC042 4C084/ZC062 4C084/ZC082 4C084/ZC312 4C084/ZC352 4C084/ZC552 4C085/KA03 4C085/KA04 4C085/KA05 4C085/LL03 4C085/LL07 4C085/LL13 4C085/LL17 4C085/LL18 4H045/AA10 4H045/AA11 4H045/AA20 4H045/AA30 4H045/BA10 4H045/BA41 4H045/CA40 4H045/DA50 4H045/DA76 4H045/EA20 4H045/EA50 4H045/FA74 4H045/GA26
优先权	60/249645 2000-11-16 US
外部链接	Espacenet

摘要(译)

本发明提供了人免疫球蛋白超家族蛋白（IGSFP）和鉴定和编码IGSFP的多核苷酸。本发明还提供了表达载体，宿主细胞，抗体，激动剂和拮抗剂。本发明还提供了用于诊断，治疗或预防与IGSFP异常表达相关的病症的方法。