(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2004-537270
(P2004-537270A)
(43)公表日　平成16年12月16日(2004.12.16)

(51)Int.Cl.[7]	F I			テーマコード（参考）
C12N 15/09	C12N	15/00	ZNAA	2G045
A61K 38/43	A61K	45/00		4B024
A61K 45/00	A61K	49/00	A	4B029
A61K 49/00	A61P	29/00		4B050
A61P 29/00	A61P	35/00		4B063

審査請求 未請求　予備審査請求 有　（全 174 頁）　最終頁に続く

(21)出願番号　特願2002-559808(P2002-559808)
(86)(22)出願日　平成13年12月13日(2001.12.13)
(85)翻訳文提出日　平成15年6月13日(2003.6.13)
(86)国際出願番号　PCT/US2001/048575
(87)国際公開番号　WO2002/059323
(87)国際公開日　平成14年8月1日(2002.8.1)
(31)優先権主張番号　60/255,963
(32)優先日　平成12年12月15日(2000.12.15)
(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　リー、アーンスティーン・エイ
アメリカ合衆国カリフォルニア州９４７０６・アルバニー・ケインズストリート　６２４

最終頁に続く

(54)【発明の名称】アミノアシルｔＲＮＡ合成酵素

(57)【要約】
本発明はヒトのアミノアシルｔＲＮＡ合成酵素（ＡＴＲＳ）および、ＡＴＲＳ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、ＡＴＲＳ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（配列番号１）のアミノ酸配列を持つポリペプチド
（ｂ）ＳＥＱ　ＩＤ　ＮＯ：１のアミノ酸配列に対して少なくとも９０％が同一であるような天然アミノ酸配列を持つポリペプチド
（ｃ）ＳＥＱ　ＩＤ　ＮＯ：１の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の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のアミノ酸配列を持つ、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：２のポリヌクレオチド配列を持つ、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持つ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持つ遺伝形質転換体。
【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、
（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されたプロモーター配列を持つ組換えポリヌクレオチドで形質転換される細胞を培養する過程と、
（ｂ）そのように発現した前記ポリペプチドを回収する過程、とからなる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のアミノ酸配列を持つことを特徴とする、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：２のポリヌクレオチド配列を持つポリヌクレオチド
（ｂ）ＳＥＱ　ＩＤ　ＮＯ：２のポリヌクレオチド配列に対して少なくとも９０％が同一であるような天然ポリヌクレオチド配列を持つポリヌクレオチド
（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群を持つ単離されたポリヌクレオチド。
【請求項１４】
請求項１２のポリヌクレオチドの配列を持つ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或る配列を有する少なくとも２０の連続したヌクレオチド群を持つ或るプローブと前記サンプルとをハイブリダイズし、該プローブが、或るハイブリダイゼーション複合体が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群との間に形成される条件下で、前記標的ポリヌクレオチドへ特異的にハイブリダイズする過程と、
（ｂ）前記ハイブリダイゼーション複合体の有無を検出し、該複合体が存在する場合にはオプションでその量を検出する過程、とを含む方法。
【請求項１５】
前記プローブが少なくとも６０の連続したヌクレオチド群を持つことを特徴とする、請求項１４に記載の方法。
【請求項１６】
請求項１２のポリヌクレオチドの配列を持つ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）前記増幅した標的ポリヌクレオチドまたはその断片の有無を検出し、該標的ポリヌクレオチドまたはその断片が存在する場合にはオプションでその量を検出する過程、とを含むことを特徴とする方法。
【請求項１７】
請求項１のポリペプチドと、薬剤として許容できる賦形剤とを有することを特徴とする組成物。
【請求項１８】
前記ポリペプチドが、ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列を持つことを特徴とする、請求項１７に記載の組成物。
【請求項１９】
機能的ＡＴＲＳの発現の低下に関連する疾患や病態の治療方法であって、そのような治療が必要な患者への請求項１７の組成物の投与を含むことを特徴とする治療方法。
【請求項２０】
請求項１のポリペプチドのアゴニストとしての有効性を確認するために化合物をスクリーニングする方法であって、
（ａ）請求項１のポリペプチドを有する或るサンプルを或る化合物に曝す過程と、
（ｂ）前記サンプルにおいてアゴニスト活性を検出する過程、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の方法で同定したアゴニスト化合物を有し、薬剤として許容できる賦形剤をも有する組成物。
【請求項２２】
機能的ＡＴＲＳの発現の低下に関連する疾患や病態の治療方法であって、そのような治療が必要な患者に請求項２１の組成物を投与することを含むことを特徴とする治療方法。
【請求項２３】
請求項１のポリペプチドのアンタゴニストとしての有効性を確認するために化合物をスクリーニングする方法であって、
（ａ）請求項１のポリペプチドを有する或るサンプルを或る化合物に曝す過程と、
（ｂ）前記サンプルにおいてアンタゴニスト活性を検出する過程、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の方法で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有する組成物。
【請求項２５】

機能的ＡＴＲＳの過剰発現に関連する疾患や病態の治療方法であって、そのような治療が必要な患者に請求項２４の組成物を投与することを含むことを特徴とする治療方法。
【請求項２６】
請求項１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）請求項１に記載のポリペプチドを適切な諸条件下で少なくとも１つの試験化合物と混合する過程と、
（ｂ）請求項１に記載のポリペプチドの試験化合物との結合を検出し、それによって請求項１に記載のポリペプチドに特異結合する化合物を同定する過程、とを含むことを特徴とする方法。
【請求項２７】
請求項１のポリペプチドの活性をモジュレート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諸条件下で、請求項１に記載のポリペプチドを少なくとも１つの試験化合物と混合する過程と、
（ｂ）請求項１に記載のポリペプチドの活性を試験化合物の存在下で算定する過程と、
（ｃ）試験化合物の存在下での請求項１に記載のポリペプチドの活性を、試験化合物の不存在下での請求項１に記載のポリペプチドの活性と比較する過程とを含み、試験化合物の存在下での請求項１に記載のポリペプチドの活性の変化が、請求項１に記載のポリペプチドの活性を調節する化合物を標示することを特徴とする方法。
【請求項２８】
請求項５の配列を持つ標的ポリヌクレオチドの発現を改変するのに効果的な化合物をスクリーニングする方法であって、
（ａ）前記標的ポリヌクレオチドの発現に好適な諸条件下で、前記標的ポリヌクレオチドを有する或るサンプルを或る化合物に曝露するステップと、
（ｂ）前記標的ポリヌクレオチドの、改変された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）核酸群を有する或る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群と、請求項１２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群を有する或る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或る標的ポリヌクレオチドとの間で或る特異的なハイブリダイゼーション複合体が形成される諸条件下で行われ、前記標的ポリヌクレオチドが、請求項１２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有する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、或る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】
生体サンプル中のＡＴＲＳの発現に関連する病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、前記生体サンプルを請求項１１に記載の抗体と混合する過程と、
（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、前記生体サンプル中の前記ポリペプチドの存在と相関することを特徴とする方法。
【請求項３１】

請求項１１の抗体であって、
（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抗体。
【請求項３２】
請求項１１の抗体と、許容できる賦形剤とを有する組成物。
【請求項３３】
被検者のＡＴＲＳ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の組成物。
【請求項３５】
被検者のＡＴＲＳ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の抗体の特異性を持つポリクローナル抗体を調製する方法であって、
（ａ）抗体応答を誘発する条件下で、ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列を持つポリペプチドまたはその免疫原性断片を用いて或る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該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のアミノ酸配列を持つ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の方法で産生したポリクローナル抗体。
【請求項３８】
請求項３７のポリクローナル抗体と好適なキャリアとを有する組成物。
【請求項３９】
請求項１１に記載の抗体の特異性を有するモノクローナル抗体を作製する方法であって、
（ａ）抗体応答を誘発する条件下で、ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列を持つポリペプチドまたはその免疫原性断片を用いて或る動物を免疫化する過程と、
（ｂ）前記動物から、抗体を産出する細胞を単離する過程と、
（ｃ）不死化した細胞と前記抗体産出細胞とを融合し、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）前記培養物からＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列を持つポリペプチドに特異結合するようなモノクローナル抗体を単離する過程、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と適切なキャリアとを有する組成物。
【請求項４２】
Ｆａｂ発現ライブラリをスクリーニングすることにより産出され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をスクリーニングすることにより産出され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のアミノ酸配列を持つポリペプチドをサンプル中に検出する方法で

あって、
（ａ）前記抗体と前記ポリペプチドとの特異結合を許容する条件下で、或るサンプルと共に請求項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、該特異結合が、ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列を持つポリペプチドがサンプル中に存在することを標示することを特徴とする方法。
【請求項４５】
或るサンプルからのＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列を持つ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との特異結合を許容する条件下で、或るサンプルと共に請求項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のアミノ酸配列を持つ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１つのエレメントが請求項１３に記載のポリヌクレオチドであることを特徴とするマイクロアレイ。
【請求項４７】
ポリヌクレオチド群を有するサンプルの発現プロフィールを作製する方法であって、
（ａ）該サンプルの該ポリヌクレオチド群を標識する過程と、
（ｂ）ハイブリダイゼーション複合体の形成に適した諸条件下で請求項４６のマイクロアレイのエレメント群をサンプルの標識されたポリヌクレオチド群と接触させる過程と、
（ｃ）該サンプル中の該ポリヌクレオチド群の発現を定量化する過程、とを含むことを特徴とする方法。
【請求項４８】
固体基板上の固有の物理的位置に付着された種々のヌクレオチド分子を有するアレイであって、少なくとも１つの前記ヌクレオチド分子が、標的ポリヌクレオチドの少なくとも３０の連続したヌクレオチド群と特異的にハイブリダイズ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有し、前記の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ことを特徴とするアレイ。
【請求項４９】
請求項４８に記載のアレイで、前記のオリゴヌクレオチドまたはポリヌクレオチドの最初の配列が前記の標的ポリヌクレオチドの少なくとも３０の連続したヌクレオチド群に完全に相補的であることを特徴とするアレイ。
【請求項５０】
請求項４８に記載のアレイで、前記のオリゴヌクレオチドまたはポリヌクレオチドの最初の配列が前記の標的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群に完全に相補的であることを特徴とするアレイ。
【請求項５１】
請求項４８に記載のアレイで、前記のオリゴヌクレオチドまたはポリヌクレオチドの最初の配列が前記の標的ポリヌクレオチドに完全に相補的であることを特徴とするアレイ。
【請求項５２】
請求項４８に記載のアレイで、マイクロアレイであることを特徴とするアレイ。
【請求項５３】
請求項４８に記載のアレイで、更に前記のオリゴヌクレオチドまたはポリヌクレオチドの最初の配列を有するヌクレオチド分子にハイブリダイズした前記の標的ポリヌクレオチドを有することを特徴とするアレイ。
【請求項５４】
請求項４８に記載のアレイで、リンカーが少なくとも１つの前記のヌクレオチド分子と前記の固体基板とを連結していることを特徴とするアレイ。
【請求項５５】

請求項４８に記載のアレイで、基板上の固有の物理的位置の各々が複数のヌクレオチド分子を持ち、任意の単一の固有の物理的位置でのその複数のヌクレオチド分子は同一の配列を有し、基板上の固有の物理的位置の各々は、基板上の別の固有の物理的位置でのヌクレオチド分子の配列とは異なる配列を有するヌクレオチド分子を持つことを特徴とするアレイ。

【請求項５６】

ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列を有する請求項１に記載のポリペプチド。

【請求項５７】

ＳＥＱ ＩＤ ＮＯ：２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、アミノアシルｔＲＮＡ合成酵素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細胞増殖異常、自己免疫／炎症性の疾患の診断・治療・予防に関する。本発明はさらに、アミノアシルｔＲＮＡ合成酵素の核酸配列及びアミノ酸配列の発現における外来化合物の効果についての算定に関する。

【０００２】

（発明の背景）

遺伝子コードの正しい翻訳は、適切な転移ＲＮＡ（ｔＲＮＡ）との連結を形成する各アミノ酸に依存する。アミノアシル－ｔＲＮＡ合成酵素（ａａＲＳ） は、全ての生物に在る本質的なタンパク質である。ａａＲＳは、タンパク質生合成の最初の段階として、アミノ酸の活性化と、その同族ｔＲＮＡとの正しい付着とを担う。原核生物は少なくとも２０種の異なるタイプのａａＲＳを持ちアミノ酸ごとに１種類のａａＲＳを持つが、真核生物は通常、異なるアミノ酸の各々のために細胞質ゾル形態とミトコンドリア形態の２つのａａＲＳを有する。２０種のａａＲＳ酵素は２つの構造的クラスに分け得る。クラスＩの酵素はｔＲＮＡの３’末端で２’ヒドロキシルにアミノ酸を加え、クラスＩＩの酵素はｔＲＮＡの３’末端で３’ヒドロキシルにアミノ酸を加える。各クラスは触媒的ドメインの特有のトポロジーによって特徴づけられる。クラスＩ酵素には、ヌクレオチド結合した「ロスマンフォールド（Ｒｏｓｓｍａｎ ｆｏｌｄ）」に基づく触媒的ドメインがある。特に、或るコンセンサステトラペプチドのモチーフは、高度に保存的である（Ｐｒｏｓｉｔｅ Ｄｏｃｕｍｅｎｔ ＰＤＯＣ００１６１， Ａｍｉｎｏａｃｙｌ－ｔｒａｎｓｆｅｒ ＲＮＡ ｓｙｎｔｈｅｔａｓｅｓ ｃｌａｓｓ－Ｉ ｓｉｇｎａｔｕｒｅ）。クラスＩ酵素は、アルギニン、システイン、グルタミン酸、グルタミン、イソロイシン、ロイシン、メチオニン、チロシン、トリプトファンおよびバリンに特異的である。クラスＩＩ酵素は、中央の触媒ドメイン（７本鎖の逆平行β－シートドメインからなる）と、Ｎ末端およびＣ末端調節ドメインを持つ。クラスＩＩ酵素は酵素のヘテロ二量体構造あるいはホモ二量体構造に基づき２群に分けられるが、後者はさらにＮ末端およびＣ末端調節ドメインの構造によって分けられる（Ｈａｒｔｌｅｉｎ， Ｍ．およびＳ．Ｃｕｓａｃｋ（１９９５）Ｊ．Ｍｏｌ．Ｅｖｏｌ．４０：５１９－５３０）。クラスＩＩ酵素は、アラニン、アスパラギン、アスパラギン酸、グリシン、ヒスチジン、リジン、フェニルアラニン、プロリン、セリンおよびトレオニンに特異的である。

【０００３】

数種のａａＲＳには編集（エディティング）機能もある。例えばＩｌｅＲＳはバリンを誤って活性化しＶａｌ－ｔＲＮＡ^{Ｉｌｅ}を形成する場合があるが、この生成物は、誤って結合した（ｍｉｓｃｈａｒｇｅｄ）生成物を破壊する或る加水分解作用によって取り除かれる。この編集活性は、クラスＩ酵素のＲｏｓｓｍａｎ ｆｏｌｄドメイン内の長い挿入配列である接続ポリペプチド １ 領域（ｃｏｎｎｅｃｔｉｖｅ ｐｏｌｙｐｅｐｔｉｄｅ １ ｒｅｇｉｏｎ）（ＣＰ１）で見つかる ２ 番目の触媒サイト内に位置する（Ｓｃｈｉｍｍｅｌ， Ｐ．他（１９９８） ＦＡＳＥＢ Ｊ．１２：１５９９－１６０９）。ａａＲ

ＳはまたｔＲＮＡプロセシングでも役割を果たす。成熟したｔＲＮＡは細胞質に移出される前に、核の中のそれぞれのアミノ酸と結合する（ｃｈａｒｇｅｄ）ことが示されている。またこのアミノ酸との結合は品質管理機構として機能し、確実にｔＲＮＡを機能的に保っている可能性がある（Ｍａｒｔｉｎｉｓ，Ｓ．Ａ．他（１９９９）ＥＭＢＯ Ｊ．１８：４５９１－４５９６）。

【０００４】

プロテイン合成における機能に加えて、特定のアミノアシルｔＲＮＡ合成酵素類は、細胞忠実度、ＲＮＡスプライシング、ＲＮＡ輸送、アポトーシスおよび、転写と翻訳の調節で役割を果たす。例えばヒトのチロシルｔＲＮＡ合成酵素はたんぱく質分解により、異なるサイトカイン活性を持つ２つの部分に切断され得る。Ｃ末端ドメインは単球と白血球の走化作用を示し、また、ミエロペルオキシダーゼ、腫瘍壊死因子－α、および組織因子の生成を刺激する。Ｎ末端ドメインはインターロイキン－８タイプＡ受容体に結合し、インターロイキン８様サイトカインとして機能する。ヒトのチロシルｔＲＮＡ合成酵素（ＴｙｒＲＳ）はアポトーシス段階の腫瘍細胞から分泌される。また、アポトーシスを加速する可能性がある（Ｗａｋａｓｕｇｉ，ＫおよびＰ．Ｓｃｈｉｍｍｅｌ（１９９９）Ｓｃｉｅｎｃｅ２８４：１４７－１５１）。ミトコンドリア型のアカパンカビ（Ｎｅｕｒｏｓｐｏｒａ ｃｒａｓｓａ）ＴｙｒＲＳおよび出芽酵母（Ｓ．ｃｅｒｅｖｉｓｉａｅ）ＬｅｕＲＳはグループＩイントロンの数種のスプライシング活性に必須であり、ヒトのミトコンドリアＬｅｕＲＳは、酵母のヌル種（ｎｕｌｌ ｓｔｒａｉｎ）において酵母ＬｅｕＲＳとして代用できる。数種のバクテリアａａＲＳは、それ自身の転写または翻訳の調節に関与している（前出Ｍａｒｔｉｎｉｓ）。数種のａａＲＳは、細胞増殖、分化、およびアポトーシスにおいて役割を持つシグナル伝達分子の１クラスである、ジアデノシン オリゴホスフェートを合成できる（Ｋｉｓｓｅｌｅｖ，Ｌ．Ｌ他（１９９８）ＦＥＢＳ Ｌｅｔｔ．４２７：１５７－１６３；Ｖａｒｔａｎｉａｎ，Ａ．他（１９９９）ＦＥＢＳ Ｌｅｔｔ．４５６：１７５－１８０）。

【０００５】

アミノアシル－ｔＲＮＡに対する自己抗体は、リウマチ関節炎、皮膚筋炎、及び多発性筋炎など、自己免疫疾患の患者によって産生される。また、複雑な間質性肺疾患（ＩＬＤ）と強度に相関する（Ｆｒｅｉｓｔ，Ｗ．他（１９９９）Ｂｉｏｌ．Ｃｈｅｍ．３８０：６２３－６４６；Ｆｒｅｉｓｔ，Ｗ．他（１９９６）Ｂｉｏｌ．Ｃｈｅｍ．Ｈｏｐｐｅ Ｓｅｙｌｅｒ ３７７：３４３－３５６）。これらの自己抗体はウイルス感染に応じて産生されるとみられ、コクサッキー ウイルスは動物に実験的ウイルス性筋炎を誘発するために用いられている。

【０００６】

ａａＲＳ構造の、人体と病原体とでの比較は、新規の抗生物質の設計に有用となっている（Ｓｃｈｉｍｍｅｌ、前出）。遺伝子操作したａａＲＳ群を用いてｉｎ ｖｉｖｏで、非天然アミノ酸群のタンパク質類への部位特異的取り込みが可能となっている（Ｌｉｕ，Ｄ．Ｒ．他（１９９７）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ ９４：１００９２－１００９７）。

【０００７】

新規のアミノアシルｔＲＮＡ合成酵素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、細胞増殖異常、自己免疫／炎症性の疾患の、診断・治療・予防において有用であり、また、アミノアシルｔＲＮＡ合成酵素の核酸配列及びアミノ酸配列の発現における外来化合物の効果についての算定にも有用である。

【０００８】

（発明の概要）

この発明は、総称して「ＡＴＲＳ」、個別には「ＡＴＲＳ－１」と呼ばれる、アミノアシルｔＲＮＡ合成酵素である精製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のアミノ酸配列からなるポリペプチド、（ｂ）ＳＥＱ Ｉ

ＤＮＯ：１のアミノ酸配列に対し少なくとも９０％が同一である天然アミノ酸配列を有するポリペプチド、（ｃ）ＳＥＱ ＩＤ ＮＯ：１の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のアミノ酸配列を有するポリペプチドの免疫原性断片、からなる群から選択した単離されたポリペプチドを提供する。一実施態様で本発明は、ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列からなる単離されたポリペプチドを提供する。

【０００９】

また本発明は、（ａ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列からなるポリペプチド、（ｂ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列に対し少なくとも９０％が同一である天然のアミノ酸配列を有するポリペプチド、（ｃ）ＳＥＱ ＩＤ ＮＯ：１の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のアミノ酸配列を有するポリペプチドの免疫原性断片、からなる群から選択された或るポリペプチドをコードする単離されたポリヌクレオチドを提供する。一実施態様では、そのポリヌクレオチドは、ＳＥＱ ＩＤ ＮＯ：１のポリペプチドをコードする。別の実施態様では、そのポリヌクレオチドはＳＥＱ ＩＤ ＮＯ：２の配列を持つ。

【００１０】

本発明は更に、（ａ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列を有するポリペプチド、（ｂ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列との少なくとも９０％の同一性を有する天然アミノ酸配列からなるポリペプチド、（ｃ）ＳＥＱ ＩＤ ＮＯ：１の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の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或るポリペプチドをコードするポリヌクレオチドと機能的に連結（リンク）したプロモーター配列を有する組換えポリヌクレオチドを提供する。一実施態様では、本発明は組換えポリヌクレオチドを用いて形質転換した細胞を提供する。別の実施態様で本発明は、この組換えポリヌクレオチドを持つ遺伝形質転換体を提供する。

【００１１】

また、本発明は（ａ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列を有するポリペプチド、（ｂ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列との少なくとも９０％の同一性を有する天然アミノ酸配列を有するポリペプチド、（ｃ）ＳＥＱ ＩＤ ＮＯ：１の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の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或るポリペプチドを製造する方法を提供する。その製造方法は、（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。

【００１２】

本発明は更に、（ａ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列を有するポリペプチド、（ｂ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列に対し少なくとも９０％同一である天然アミノ酸配列を有するポリペプチド、（ｃ）ＳＥＱ ＩＤ ＮＯ：１の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の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或る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。

【００１３】

本発明は更に、（ａ）ＳＥＱ ＩＤ ＮＯ：２の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：２のポリヌクレオチド配列との少なくとも９０％の同一性を有する天然ポリヌクレオチド配列を有するポリヌクレオチド、（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、または（ｅ）（ａ）～（ｄ）のＲＮＡ等価物、を含む単離されたポリヌクレオチドを提供する。一実施態様で該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群からなる。

【００１４】

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：２の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：２のポリヌクレオチド配列との少なくとも９０％の同一性を有する天然ポリヌクレオチド配列を有するポリヌクレオチド、（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、または（ｅ）（ａ）～（ｄ）のＲＮＡ等価物、を含む群から選択された或るポリヌクレオチドの配列を有する。検出方法は、（ａ）サンプル中の上記標的ポリヌクレオチドに相補的な或る配列からなる少なくとも２０の連続したヌクレオチド群からなる或るプローブを用いて該サンプルをハイブリダイズする過程と、（ｂ）該ハイブリダイゼーション複合体の有無を検出し、複合体が存在すればオプションでその量を検出する過程からなる。該プローブと該標的ポリヌクレオチドあるいはその断片との間でハイブリダイゼーション複合体が形成されるような条件下で、プローブは、該標的ポリヌクレオチドに対し特異的にハイブリダイズする。一実施態様では、プローブは少なくとも６０の連続したヌクレオチド群からなる。

【００１５】

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：２の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：２のポリヌクレオチド配列との少なくとも９０％の同一性を有する天然ポリヌクレオチド配列を有するポリヌクレオチド、（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、または（ｅ）（ａ）～（ｄ）のＲＮＡ等価物、を含む群から選択された或るポリヌクレオチドの配列を有する。検出方法は、（ａ）ポリメラーゼ連鎖反応増幅を用いて標的ポリヌクレオチドまたはその断片を増幅する過程と、（ｂ）増幅した標的ポリヌクレオチドまたはその断片の有無を検出し、該標的ポリヌクレオチドまたはその断片が存在すればオプションでその量を検出する過程からなる。

【００１６】

本発明は更に、有効量のポリペプチドと、薬剤として許容できる賦形剤とからなる組成物を提供する。有効量のポリペプチドは、（ａ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列を有するポリペプチド、（ｂ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列との少なくとも９０％の同一性を有する天然アミノ酸配列を含むポリペプチド、（ｃ）ＳＥＱ ＩＤ ＮＯ：１の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のアミノ酸配列を有するポリペプチドの免疫原性断片、からなる群から選択される。一実施態様では、この組成物はＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列を持つ。更に本発明は、この組成物を機能的ＡＴＲＳの発現の低下に関連した疾患や症状の治療を必要とする患者に投与する方法を提供する。

【００１７】

本発明はまた、（ａ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列を有するポリペプチド、（ｂ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列との少なくとも９０％の同一性を有する天然アミノ酸配列を有するポリペプチド、（ｃ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列の生物学的活性断片、または（ｄ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列の免疫原性断片、を含む群から選択された或る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、この方法で同定したアゴニスト化合物と許容される医薬用賦形剤とを有する、或る組成物を提供する。更なる別法では、本発明は、この組成物を機能的ＡＴＲＳの発現の低下に関連した疾患やその症状の治療を必要とする患者に投与する方法を提供する。

【００１８】

本発明は更に、（ａ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列を有するポリペプチド、（ｂ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列との少なくとも９０％の同一性を有する天然アミ

ノ酸配列を有するポリペプチド、（ｃ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列の生物学的活性断片、または（ｄ）ＳＥＱ ＩＤ ＮＯ：１の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）サンプル中のアンタゴニスト活性を検出する過程とを含む。別の一実施態様で本発明は、この方法で同定したアンタゴニスト化合物と薬剤として許容し得る賦形剤とを有する組成物を提供する。更なる別法では、本発明は、機能的ＡＴＲＳの過剰な発現に関連した疾患や症状の治療を必要とする患者への、この組成物の投与を含む方法を提供する。

【００１９】

本発明は更に、（ａ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列を持つポリペプチド、（ｂ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列との少なくとも９０％の同一性を持つ天然アミノ酸配列を有するポリペプチド、（ｃ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列を持つポリペプチドの生物学的活性断片、または（ｄ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列を持つポリペプチドの免疫原性断片、を含む群から選択された或る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）該ポリペプチドを適切な条件下で少なくとも１つの試験化合物と混合する過程と、（ｂ）試験化合物との該ポリペプチドの結合を検出し、それによって該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。

【００２０】

本発明は更に、（ａ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列を有するポリペプチド、（ｂ）ＳＥＱ ＩＤ ＮＯ：１のアミノ酸配列との少なくとも９０％の同一性を有する天然アミノ酸配列を有するポリペプチド、（ｃ）ＳＥＱ ＩＤ ＮＯ：１の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の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或る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）該ポリペプチドの活性が許容された条件下で、該ポリペプチドを少なくとも１つの試験化合物と混合する過程と、（ｂ）該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該ポリペプチドの活性を試験化合物の不存在下での該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該ポリペプチドの活性の変化は、該ポリペプチドの活性を調節する化合物であることを標示する。

【００２１】

更に本発明は、ＳＥＱ ＩＤ ＮＯ：２のポリヌクレオチド配列を持つ標的ポリヌクレオチドの発現を改変する効果につき、或る化合物をスクリーニングする一方法を提供する。この方法は、（ａ）この標的ポリヌクレオチドを有するサンプルを或る化合物に曝露する過程と、（ｂ）この標的ポリヌクレオチドの発現の改変を検出する過程と、（ｃ）可変量のこの化合物の存在下でのこの標的ポリヌクレオチドの発現と、この化合物の不在下での発現とを比較する過程とからなる。

【００２２】

本発明は更に、（ａ）核酸群を有する或る生体サンプルを試験化合物で処理する過程と、（ｂ）（ｉ）ＳＥＱ ＩＤ ＮＯ：２のポリヌクレオチド配列からなるポリヌクレオチド、（ｉｉ）ＳＥＱ ＩＤ ＮＯ：２のポリヌクレオチド配列との少なくとも９０％の同一性を有する天然ポリヌクレオチド配列からなるポリヌクレオチド、（ｉｉｉ）（ｉ）に相補的な配列を有するポリヌクレオチド、（ｉｖ）（ｉｉ）のポリヌクレオチドに相補的なポリヌクレオチド、（ｖ）（ｉ）～（ｉｖ）のＲＮＡ等価物、を含む群から選択した或る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群から構成される或るプローブを用いて、処理した生体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、上記プローブと生体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：２のポリヌク

レオチド配列からなるポリヌクレオチド、（ｉｉ）ＳＥＱ ＩＤ ＮＯ：２のポリヌクレオチド配列との少なくとも９０％の同一性を有する天然ポリヌクレオチド配列からなるポリヌクレオチド、（ｉｉｉ）（ｉ）のポリヌクレオチドに相補的な配列を有するポリヌクレオチド、（ｉｖ）（ｉｉ）のポリヌクレオチドに相補的なポリヌクレオチド、（ｖ）（ｉ）～（ｉｖ）のＲＮＡ等価物、を含む群から選択する。あるいは、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を持つ。毒性の算定方法には更に、以下の過程がある。（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）処理済み生体サンプル中のハイブリダイゼーション複合体の量を、非処理の生体サンプル中のハイブリダイゼーション複合体の量と比較する過程である。処理済み生体サンプル中のハイブリダイゼーション複合体の量の差異が、試験化合物の毒性を示す。

【００２３】

（本発明の記述について）

本発明のタンパク質、ヌクレオチド配列および方法について説明するが、その前に、説明した特定の装置、材料および方法に本発明が限定されるものではなく、修正され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態様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。

【００２４】

請求の範囲および明細書中で用いている単数形の「或る」および「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されたい。したがって、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、および、当業者に公知の抗体の等価物などについても言及している。

【００２５】

本明細書中で用いる全ての技術用語および科学用語は、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、材料および方法を用いて本発明の実施または試験を行うことができるが、ここでは好適な装置、材料、方法について説明する。本発明で言及する全ての刊行物は、刊行物中で報告されていて且つ本発明に関係があるであろう細胞、プロトコル、試薬およびベクターについて説明およ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。

【００２６】

（定義）

用語「ＡＴＲＳ」は、天然、合成、半合成或いは組換え体などの、全ての種（特にウシ、ヒツジ、ブタ、ネズミ、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ＡＴＲＳのアミノ酸配列を指す。

【００２７】

用語「アゴニスト」は、ＡＴＲＳの生物学的活性を強めたり、模倣する分子を指す。アゴニストとしては、ＡＴＲＳと直接相互作用して、或いはＡＴＲＳが関与する生物学的経路の諸成分に作用してＡＴＲＳ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。

【００２８】

用語「対立遺伝子変異配列」は、ＡＴＲＳをコードする別の形の遺伝子を指す。対立遺伝子変異配列群は、核酸配列における少なくとも１つの突然変異から生じ得る他、変容したｍＲＮＡ群を生じ得る。また、ポリペプチド群を生じ得るが、それらのポリペプチドの構造または機能は、変容することもしないこともある。或る遺伝子は、その天然型の対立遺伝子変異配列を全く持たない場合もあり、１個以上持つこともある。対立遺伝子変異配列を生じさせる通常の突然変異性変化は一般に、ヌクレオチドの自然な欠失、付加または置

換による。これら各種の変化は、単独であるいは他の変化と共に、或る配列内で１回以上、生じ得る。ＡＴＲＳをコードする「変容した／改変された」核酸配列としては、様々なヌクレオチドの欠失、挿入、或いは置換がありながら、ＡＴＲＳと同じポリペプチド、或いはＡＴＲＳの機能特性の少なくとも１つを備えるポリペプチドをもたらす配列もある。この定義には、ＡＴＲＳをコードするポリヌクレオチド配列にとり正常な染色体の遺伝子座ではない位置でのアレル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を含み、また、ＡＴＲＳをコードするポリヌクレオチドの或る特定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も含む。コードされるタンパク質も「変容する／改変される」ことがあり、また、サイレント変化を生じて機能的には等価なＡＴＲＳとなるような、アミノ酸残基の欠失、挿入、或いは置換を持ち得る。意図的なアミノ酸置換は、生物学的或いは免疫学的にＡＴＲＳ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およ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およびアルギニンがある。親水性値が近似した非荷電極性側鎖を持つアミノ酸には、アスパラギンとグルタミン、セリンとトレオニンがある。親水性値が近似した非荷電側鎖を持つアミノ酸には、ロイシンとイソロイシンとバリン、グリシンとアラニン、フェニルアラニンとチロシンがある。

【００２９】

用語「アミノ酸」および「アミノ酸配列」は、オリゴペプチド、ペプチド、ポリペプチド、タンパク質配列、あるいはそれらのいずれかの断片を指し、天然分子または合成分子を指す。「アミノ酸配列」が或る天然タンパク質分子の配列を指す場合、「アミノ酸配列」および類似の用語は、記載したそのタンパク質分子に関連する完全で元のままのアミノ酸配列に限定するものではない。

【００３０】

用語「増幅」は、或る核酸配列の付加的複製物を作製する行為に関する。増幅には通常、当業者に公知のポリメラーゼ連鎖反応（ＰＣＲ）技術を用いる。

【００３１】

用語「アンタゴニスト」は、ＡＴＲＳの生物学的活性を阻害或いは減弱する分子を指す。アンタゴニストとしては、ＡＴＲＳと直接相互作用して、或いはＡＴＲＳが関与する生物学的経路の諸成分に作用してＡＴＲＳの活性を調節する、タンパク質（抗体など）、核酸、糖質、小分子、任意の他の化合物や組成物などを含み得る。

【００３２】

用語「抗体」は、エピトープの決定基と結合できる、無傷の免疫グロブリン分子やそれらの断片、例えばＦａｂ、Ｆ（ａｂ’）$_2$、およびＦｖ断片を指す。ＡＴＲＳポリペプチド群と結合する抗体類は、免疫抗原として、無傷のポリペプチド群、または目的の小ペプチド群を有する断片群を用いて調製できる。マウス、ラット、ウサギなどの動物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の由来は、ＲＮＡの翻訳の場合や、または化学合成があり得る。また、それらは所望により、キャリアタンパク質に抱合し得る。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キャリアの例は、ウシ血清アルブミン、サイログロブリン、および、スカシガイのヘモシアニン（ＫＬＨ）などがある。結合したこのペプチドは、前記動物を免疫化するために用いる。

【００３３】

用語「抗原決定基」は、特定の抗体と接触する、分子の領域（すなわちエピトープ）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発し得る。或る抗原決定基は、或る抗体への結合について、無傷抗原（すなわち免疫応答を引き出すために用いられる免疫原）と競合し得る。

【００３４】

用語「アプタマー（ａｐｔａｍｅｒ）」は、核酸またはオリゴヌクレオチド分子であって、特定の分子ターゲットに結合する分子を指す。アプタマーはｉｎ ｖｉｔｒｏでの進化

プロセスに由来する（例えば、ＳＥＬＥＸ（Ｓｙｓｔｅｍａｔｉｃ Ｅｖｏｌｕｔｉｏｎ ｏｆ Ｌｉｇａｎｄｓ ｂｙ ＥＸｐｏｎｅｎｔｉａｌ Ｅｎｒｉｃｈｍｅｎｔの略、試験管内選択法）、米国特許第５，２７０，１６３号に記述）。これは、大規模な組み合わせライブラリ群から標的特異的アプタマー配列を選択するプロセスである。アプタマーの構成は二本鎖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または他のヌクレオチド様分子を含み得る。アプタマーの各ヌクレオチド成分は修飾された糖基を持つことがあり（例えば、リボヌクレオチドの２’－ＯＨ基は２’－Ｆまたは２’－NH_2によって置換され得る）、これは、ヌクレアーゼへの抵抗性または血中でのより長い寿命など、所望の特性を向上し得る。アプタマーを他の分子（例えば高分子量のキャリア）と抱合させることにより、循環系からのこのアプタマーのクリアランスを遅らせ得る。アプタマー類は、例えば或る架橋剤の光活性化によって、それらの同種リガンド群と特異的に架橋させ得る（例えば、Ｂｒｏｄｙ，Ｅ．Ｎ．およびＬ．Ｇｏｌｄ（２０００）Ｊ．Ｂｉｏｔｅｃｈｎｏｌ．７４：５－１３を参照）。

【００３５】

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に基づく或るＲＮＡ発現系を用いて、白血球の細胞質内で特定のＲＮＡアプタマー類が高レベルに発現されている（Ｂｌｉｎｄ，Ｍ．他（１９９９）Ｐｒｏｃ．Ｎａｔｌ Ａｃａｄ．Ｓｃｉ．ＵＳＡ ９６：３６０６－３６１０）。

【００３６】

用語「スピーゲルマー（ｓｐｉｅｇｅｌｍｅｒ）」は、アプタマーの内、Ｌ－ＤＮＡ、Ｌ－ＲＮＡなどの左旋性ヌクレオチド誘導体または左旋性ヌクレオチド様分子などを指す。左旋性のヌクレオチド群を持つアプタマー類は、右旋性ヌクレオチド群を持つ基質に通常は作用する、天然酵素類による分解に対して耐性がある。

【００３７】

用語「アンチセンス」は、或る特定の核酸配列の「センス」（コーディング）鎖との塩基対を形成し得る任意の組成物を指す。アンチセンス組成物としては、ＤＮＡや、ＲＮＡや、ペプチド核酸（ＰＮＡ）や、修飾されたバックボーン連結たとえばホスホロチオ酸、メチルホスホン酸またはベンジルホスホン酸などを有するオリゴヌクレオチドや、修飾された糖基たとえば２’－メトキシエチル糖または２’－メトキシエトキシ糖などを有するオリゴヌクレオチドや、あるいは修飾された塩基たとえば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などを有するオリゴヌクレオチドがあり得る。アンチセンス分子は、化学合成または転写など、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、細胞に導入されると、細胞が産生した天然核酸配列との塩基対を形成し、二重鎖を形成して転写または翻訳を妨害する。「負」または「マイナス」という表現は、或る参考ＤＮＡ分子のアンチセンス鎖を指すことがあり、「正」または「プラス」という表現は、或る参考ＤＮＡ分子のセンス鎖を意味し得る。

【００３８】

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

【００３９】

用語「相補的」は、塩基対合によってアニールする２つの一本鎖核酸配列間の関係を指す。例えば、「５’－ＡＧＴ－３’」は、その相補配列「３’－ＴＣＡ－５’」との対を形成する。

【００４０】

「～のポリヌクレオチド配列を含む（持つ）組成物」または「～のアミノ酸配列を含む（持つ）組成物」は、広い意味で、指定のポリヌクレオチド配列若しくはアミノ酸配列を持つ、任意の組成物を指す。この組成物には、乾燥製剤または水溶液が含まれ得る。ＡＴＲ

Ｓをコード若しくはＡＴＲＳの断片群をコードするポリヌクレオチド配列群を有する組成物は、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も可能である。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）およびその他の成分（例えばデンハート液、粉乳、サケの精子のＤＮＡなど）を含む水溶液中にプローブを分散させることができる。

【００４１】

「コンセンサス配列」は、不要な塩基を除くためにＤＮＡ配列解析を繰り返し行い、ＸＬ－ＰＣＲキット（Ａｐｐｌｉｅｄ Ｂｉｏｓｙｓｔｅｍｓ，Ｆｏｓｔｅｒ Ｃｉｔｙ ＣＡ）を用いて５’および／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）またはＰｈｒａｐ（Ｕｎｉｖｅｒｓｉｔｙ ｏｆ Ｗａｓｈｉｎｇｔｏｎ，Ｓｅａｔｔｌｅ ＷＡ）などの断片構築用コンピュータプログラムを用いて１つあるいはそれ以上のオーバーラップするｃＤＮＡやＥＳＴ、またはゲノムＤＮＡ断片から構築された核酸配列を指す。伸長および構築の両方を行ってコンセンサス配列を産生する配列もある。

【００４２】

用語「保存的なアミノ酸置換」は、元のタンパク質の特性を殆ど変えないと予測される置換を指す。すなわち、そのタンパク質の構造、特にその機能が保存され、置換によって大きくは変わらない。下表は、タンパク質内で元のアミノ酸と置換され得るアミノ酸であり、保存的アミノ酸置換と認められるアミノ酸を示す。

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

【００４３】
保存的な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螺旋構造、（ｂ）置換部位における分子の電荷または疎水性、および／または（ｃ）側鎖の大部分、を保持する。
【００４４】
用語「欠失」は、１個以上のアミノ酸残基が欠如するアミノ酸配列内の変化、あるいは１個以上のヌクレオチドが欠如するヌクレオチド配列内の変化を指す。
【００４５】
用語「誘導体」は、化学修飾されたポリヌクレオチドまたはポリペプチドを指す。例えば、アルキル基、アシル基、ヒドロキシル基またはアミノ基による水素の置換は、ポリヌクレオチド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、次のようなプロセスによって修飾されたポリペプチドである。すなわち、グリコシル化、ポリエチレングリコール化（ｐｅｇｙｌａｔｉｏｎ）、あるいは任意の同様なプロセスであって、誘導元のポリペプチドの少なくとも１つの生物学的若しくは免疫学的機能を保持するプロセスである。

【００４６】
「検出可能な標識」は、測定可能な信号を発生し得る、ポリヌクレオチドやポリペプチドに共有結合あるいは非共有結合するレポーター分子や酵素を指す。
【００４７】
「示差発現」は、少なくとも２例のサンプルを比較して判定する、増加（上方調節）、あるいは減少（下方調節）、または欠損した、遺伝子またはタンパク質の発現を指す。このような比較は例えば、処理済サンプルと不処理サンプル、または病態サンプルと健常サンプルとの間で行われ得る。
【００４８】
「エキソンシャッフリング」は、異なるコード領域（エキソン）群の組換えを意味する。或るエキソンは、コードするタンパク質の１つの構造的または機能的ドメインを提示し得るため、安定したサブストラクチャー群の新たな組み合わせによって新規のタンパク質群を構築することが可能であり、新たなタンパク質機能の進化を促進できる。
【００４９】
用語「断片」は、ＡＴＲＳまたはＡＴＲＳ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、親配列より長さが短いものを指す。断片は、定義された配列の全長から１ヌクレオチド／アミノ酸残基を差し引いた長さよりも短い長さを有し得る。例えば或る断片は、５～１０００の連続したヌクレオチドまたはアミノ酸残基を有し得る。プローブ、プライマー、抗原、治療用分子として、あるいはその他の目的のために用いられる或る断片は、少なくとも長さ５、１０、１５、２０、２５、３０、４０、５０、６０、７５、１００、１５０、２５０若しくは５００の、連続したヌクレオチドあるいはアミノ酸残基であり得る。断片は、或る分子の特定領域から優先的に選択し得る。例えば、或るポリペプチド断片は、定義された或る配列内に見られるような或るポリペプチドの最初の２５０または５００アミノ酸（または最初の２５％または５０％）から選択した、或る長さの連続したアミノ酸を持ち得る。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、本発明の実施態様では、配列表、表および図面を含む本明細書が支持する任意の長さであり得る。
【００５０】
ＳＥＱ　ＩＤ　ＮＯ：２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：２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持つ。ＳＥＱ　ＩＤ　ＮＯ：２のある断片は、例えば、ハイブリダイゼーションや増幅技術、またはＳＥＱ　ＩＤ　ＮＯ：２を関連ポリヌクレオチド配列から区別する類似の方法に有用である。ＳＥＱ　ＩＤ　ＮＯ：２の或る断片の正確な長さ、及び該断片が対応するＳＥＱ　ＩＤ　ＮＯ：２の領域は、その断片に対し意図した目的に基づき当業者が慣例的に判定できる。
【００５１】
ＳＥＱ　ＩＤ　ＮＯ：１の或る断片は、ＳＥＱ　ＩＤ　ＮＯ：２の或る断片によってコードされる。ＳＥＱ　ＩＤ　ＮＯ：１の或る断片は、ＳＥＱ　ＩＤ　ＮＯ：１を特異的に同定する固有のアミノ酸配列の領域を持つ。例えば、ＳＥＱ　ＩＤ　ＮＯ：１の或る断片は、ＳＥＱ　ＩＤ　ＮＯ：１を特異的に認識する抗体の開発における免疫原性ペプチドとして有用である。ＳＥＱ　ＩＤ　ＮＯ：１の或る断片の正確な長さは、また、この断片に対応するＳＥＱ　ＩＤ　ＮＯ：１の領域は、その断片の目的に基づいて当業者が慣例的に判定し得る。
【００５２】
「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）と、それに続く１オープンリーディングフレームおよび翻訳終止コドンを有する配列である。或る「完全長」ポリヌクレオチド配列は、或る「完全長」ポリペプチド配列をコードする。
【００５３】
「相同性」の語は、２つ以上のポリヌクレオチド配列または２つ以上のポリペプチド配列の、配列類似性、互換性、または配列同一性を意味する。
【００５４】

ポリヌクレオチド配列に適用される「一致率」または「％一致」の語は、標準化されたアルゴリズムを用いてアラインメントされた、２つ以上のポリヌクレオチド配列間で一致する残基の割合を意味する。標準化アルゴリズムは、２配列間のアラインメントを最適化するため、標準化された再現性のある方法で比較対象の２配列内にギャップ群を挿入し得るので、２つの配列をより有意に比較できる。

【００５５】

ポリヌクレオチド配列間の一致率を判定するには、ＭＥＧＡＬＩＧＮ ｖｅｒｓｉｏｎ ３．１２ｅ配列アラインメントプログラムなどに組込まれているＣＬＵＳＴＡＬ Ｖアルゴリズムの、デフォルトのパラメータ群を用い得る。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．およびＰ．Ｍ． Ｓｈａｒｐ（１９８９）ＣＡＢＩＯＳ ５：１５１－１５３、Ｈｉｇｇｉｎｓ，Ｄ．Ｇ． 他（１９９２）ＣＡＢＩＯＳ ８：１８９－１９１に記載されている。ポリヌクレオチド配列をペアワイズアラインメントする際のデフォルトパラメータは、Ｋｔｕｐｌｅ＝２、ｇａｐ ｐｅｎａｌｔｙ＝５、ｗｉｎｄｏｗ＝４、「ｄｉａｇｏｎａｌｓ ｓａｖｅｄ」＝４と設定される。「ｗｅｉｇｈｔｅｄ」残基重み付け表がデフォルトで選択される。ＣＬＵＳＴＡＬ Ｖによる一致率の報告は、アラインメントされたポリヌクレオチド配列間の「ｐｅｒｃｅｎｔ ｓｉｍｉｌａｒｉｔｙ（類似性パーセント）」としてなされる。

【００５６】

あるいは、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、無料で利用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）。ＢＬＡＳＴ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およびインターネット（ｈｔｔｐ：／／ｗｗｗ．ｎｃｂｉ．ｎｌｍ．ｎｉｈ．ｇｏｖ／ＢＬＡＳＴ／）からも入手可能である。このＢＬＡＳＴソフトウェア一式には、既知のポリヌクレオチド配列と、様々なデータベースの別のポリヌクレオチド配列とのアラインメントに用いる「ｂｌａｓｔｎ」を含む、様々な配列分析プログラムが含まれる。「ＢＬＡＳＴ ２ Ｓｅｑｕｅｎｃｅｓ」と呼ばれるツールも入手可能であり、これは２つのヌクレオチド配列を直接のペアワイズで比較するために用いられる。「ＢＬＡＳＴ ２ Ｓｅｑｕｅｎｃｅｓ」は、ｈｔｔｐ：／／ｗｗｗ．ｎｃｂｉ．ｎｌｍ．ｎｉｈ．ｇｏｖ／ｇｏｒｆ／ｂｌ２．ｈｔｍｌにアクセスして、対話形式で利用できる。「ＢＬＡＳＴ ２ Ｓｅｑｕｅｎｃｅｓ」ツールは、ｂｌａｓｔｎ および ｂｌａｓｔｐ（以下に記載）の両方に用い得る。ＢＬＡＳＴプログラムは、一般的には、ギャップ（ｇａｐ）などのパラメータをデフォルト設定にセットして用いる。例えば、２つのヌクレオチド配列を比較するには、デフォルトパラメータに設定した「ＢＬＡＳＴ ２ Ｓｅｑｕｅｎｃｅｓ」ツールＶｅｒｓｉｏｎ ２．０．１２（２０００年４月２１日）を用いてｂｌａｓｔｎを実行し得る。デフォルトパラメータの設定例を以下に示す。

【００５７】

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

一致率は、ある定義された配列の全長（例えば特定のＳＥＱ ＩＤナンバーで定義された配列）で測定し得る。あるいは、より短い長さ、例えば、定義された、より大きな配列か

ら得た断片（例えば少なくとも２０、３０、４０、５０、７０、１００または少なくとも２００の連続したヌクレオチドの断片）の長さの一致率も測定し得る。ここに挙げた長さは単なる例示的なものに過ぎず、表、図および配列リストを含めた本明細書に記載された配列が支持する任意の断片長を用いて、一致率を測定し得る或る長さを説明し得ることを理解されたい。

【００５８】

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。

【００５９】

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

【００６０】

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群を用いて判定できる（参照先と共に上述）。ＣＬＵＳＴＡＬ Ｖを用いて、ポリペプチド配列をペアワイズアラインメントする際のデフォルトパラメータは、Ｋｔｕｐｌｅ＝１、ｇａｐ ｐｅｎａｌｔｙ＝３、ｗｉｎｄｏｗ＝５、「ｄｉａｇｏｎａｌｓ ｓａｖｅｄ」＝５と設定される。ＰＡＭ２５０マトリクスが、デフォルトの残基重み付け表として選択される。ポリヌクレオチドのアラインメントと同様に、アラインメントされたポリペプチド配列の対の一致率は、ＣＬＵＳＴＡＬ Ｖによって「ｐｅｒｃｅｎｔ ｓｉｍｉｌａｒｉｔｙ（類似性パーセント）」として報告される。

【００６１】

あるいは、ＮＣＢＩ ＢＬＡＳＴソフトウェア一式を用い得る。例えば、２つのポリペプチド配列をペアワイズで比較する場合、「ＢＬＡＳＴ ２ Ｓｅｑｕｅｎｃｅｓ」ツールＶｅｒｓｉｏｎ ２．０．１２（２０００年４月２１日）のｂｌａｓｔｐをデフォルトパラメータに設定して用い得る。デフォルトパラメータの設定例を以下に示す。

【００６２】

Matrix: BLOSUM62

Open Gap: 11 及び Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

一致率は、ある定義されたポリペプチド配列の全長（例えば特定のＳＥＱ ＩＤナンバーで定義された配列）で測定し得る。あるいは、より短い長さ、例えば、定義された、より大きなポリペプチド配列から得た断片（例えば少なくとも１５、２０、３０、４０、５０、７０、または少なくとも１５０の連続した残基の断片）の長さの一致率も測定し得る。ここに挙げた長さは単なる例示的なものに過ぎず、表、図および配列リストを含めた本明

細書に記載された配列が支持する任意の断片長を用いて、一致率を測定し得る或る長さを説明し得ることを理解されたい。

【００６３】

「ヒト人工染色体（ＨＡＣ）」は直鎖状の微小染色体であり、約６ｋｂ ～１０ＭｂのサイズのＤＮＡ配列を持ち得る。また、染色体の複製、分離および維持に必要な、全てのエレメントを持つ。

【００６４】

用語「ヒト化抗体」は、もとの結合能力を保持しつつ、よりヒトの抗体に似せるために、非抗原結合領域のアミノ酸配列を改変した抗体分子を指す。

【００６５】

「ハイブリダイゼーション」とは、所定のハイブリダイゼーション条件下で、或る一本鎖ポリヌクレオチドが或る相補的な一本鎖と塩基対を形成するアニーリングのプロセスである。特異的ハイブリダイゼーションは、２つの核酸配列が高い相補性を共有することの指標である。特異的ハイブリダイゼーション複合体は許容されるアニーリング条件下で形成され、１回以上の「洗浄」ステップ後もハイブリダイズされたままである。洗浄ステップは、ハイブリダイゼーションプロセスのストリンジェンシーを決定する際に特に重要であり、よりストリンジェントな条件では、非特異結合（すなわち完全には一致しない核酸鎖対間の結合）が減少する。核酸配列のアニーリングに関する許容条件は、当業者が慣例的に決定できる。アニール条件はどのハイブリダイゼーション実験でも一定であり得るが、洗浄条件は、所望のストリンジェンシー、したがってハイブリダイゼーション特異性を得るように、実験ごとに変更し得る。許容されるアニーリング条件は、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の存在下である。

【００６６】

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、所定のイオン強度およびｐＨにおける特定配列の融点（Ｔｍ）より約５～２０℃低くなるように選択する。このＴｍは、所定のイオン強度およびｐＨの条件下で、完全一致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を計算する式および核酸のハイブリダイゼーション条件はよく知られており、Ｓａｍｂｒｏｏｋ，Ｊ．他（１９８９）Ｍｏｌｅｃｕｌａｒ Ｃｌｏｎｉｎｇ：Ａ Ｌａｂｏｒａｔｏｒｙ Ｍａｎｕａｌ，第２版，１－３巻，Ｃｏｌｄ Ｓｐｒｉｎｇ Ｈａｒｂｏｒ Ｐｒｅｓｓ，Ｐｌａｉｎｖｉｅｗ ＮＹに記載されており、特に２巻の９章を参照されたい。

【００６７】

本発明のポリヌクレオチド間の高ストリンジェンシー条件のハイブリダイゼーションには、約０．２× ＳＳＣおよび約０．１％のＳＤＳの存在下、６８℃で１時間の洗浄条件を含む。別法では、温度は約６５℃、６０℃、５５℃、または４２℃を用い得る。ＳＳＣ濃度は、約０．１％のＳＤＳ存在下で、約０．１～２×ＳＳＣの範囲で変更し得る。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、せん断した変性サケ精子ＤＮＡがある。特定条件下で、例えばＲＮＡとＤＮＡのハイブリダイゼーションでは、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。進化的類似性は、ヌクレオチド群、およびヌクレオチドがコードするポリペプチド群について、或る同様の役割を強く示唆する。

【００６８】

用語「ハイブリダイゼーション複合体」は、相補的な塩基間の水素結合の形成によって形成された、２つの核酸配列の複合体を指す。ハイブリダイゼーション複合体は、溶解状態で形成し得る（C_0tまたはR_0t解析など）。あるいは、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、ある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。

【００６９】

用語「挿入」あるいは「付加」は、１個以上のアミノ酸残基あるいはヌクレオチドがそれぞれ追加される、アミノ酸配列あるいは核酸配列における変化を指す。

【００７０】

「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などに関連する症状を指し得る。これらの症状は、細胞およ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得る。

【００７１】

「免疫原性断片」は、例えば哺乳類など生物に導入すると免疫応答を引き起こす、ＡＴＲＳのポリペプチド断片またはオリゴペプチド断片である。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、ＡＴＲＳのあらゆるポリペプチド断片またはオリゴペプチド断片を含む。

【００７２】

用語「マイクロアレイ」は、或る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。

【００７３】

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

【００７４】

用語「モジュレート」または「活性を調節」は、ＡＴＲＳの活性を変化させることを指す。例えば、モジュレートによって、ＡＴＲＳのタンパク質活性の増減、或いは、結合特性、またはその他の生物学的特性、機能的特性或いは免疫学的特性の変化を生じ得る。

【００７５】

「核酸」および「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」および「核酸配列」の語はまた、ゲノム起源または合成起源のＤＮＡまたはＲＮＡであって一本鎖または二本鎖であるかあるいはセンス鎖またはアンチセンス鎖を提示し得るようなＤＮＡまたはＲＮＡや、ペプチド核酸（ＰＮＡ）や、任意のＤＮＡ様またはＲＮＡ様物質を指すこともある。

【００７６】

「機能的に連結した」は、第１の核酸配列と第２の核酸配列が機能的な関係にある状態を指す。例えば、或るプロモーターが或る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配列群は非常に近接するか連続的に隣接することがあり、また、２つのタンパク質コード領域を結合するために必要な場合は同一リーディングフレーム内にあり得る。

【００７７】

「ペプチド核酸（ＰＮＡ）」は、末端がリシンで終わるアミノ酸残基のペプチドのバックボーンに結合した、少なくとも約５ヌクレオチドの長さのオリゴヌクレオチドを持つ、アンチセンス分子または抗遺伝子剤を指す。末端のリシンは、この組成に溶解性を与える。ＰＮＡは、相補的一本鎖ＤＮＡまたはＲＮＡに優先的に結合して転写の伸長を停止させる。ポリエチレングリコール化することにより、細胞におけるＰＮＡの寿命を延長し得る。

【００７８】

ＡＴＲＳ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、蛋白分解的切断その他の、当分野で既知の修飾を含み得る。これらのプロセスは、合成或いは生化学的に生じ得る。生化学的修飾は、ＡＴＲＳの酵素環境に依存し、細胞の種類によって異なることとなる。

【００７９】

「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、ＡＴＲＳやそれらの相補配列群を、またはそれらの断片群をコードする核酸配列群を指す。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配列である。典型的な標識には、放射性アイソトープ、リガンド、化学発光試薬および酵素がある。「プライマー」とは、相補的な塩基対を形成して標的ポリヌクレオチドにアニーリング可能な、通常はＤＮＡオリゴヌクレオチドである短い核酸である。プライマーは次に、ＤＮＡポリメラーゼ酵素により、標的ＤＮＡ鎖に沿って伸長され得る。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（および同定）に用い得る。

【００８０】

本発明に用いるプローブおよ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群からなる。特異性を高めるため、長めのプローブおよびプライマー、例えば開示した核酸配列の少なくとも２０、２５、３０、４０、５０、６０、７０、８０、９０、１００または少なくとも１５０の連続したヌクレオチドからなるようなプローブおよびプライマーも用い得る。これよりもかなり長いプローブおよびプライマーもある。表、図面および配列リストを含む本明細書が支持する、任意の長さのヌクレオチドを用い得るものと理解されたい。

【００８１】

プローブおよびプライマーの調製および使用方法については、Ｓａｍｂｒｏｏｋ，Ｊ．他（１９８９）Ｍｏｌｅｃｕｌａｒ Ｃｌｏｎｉｎｇ：Ａ Ｌａｂｏｒａｔｏｒｙ Ｍａｎｕａｌ，第２版，１－３巻，Ｃｏｌｄ Ｓｐｒｉｎｇ Ｈａｒｂｏｒ Ｐｒｅｓｓ，Ｐｌａｉｎｖｉｅｗ ＮＹ、Ａｕｓｕｂｅｌ，Ｆ．Ｍ．他（１９８７）Ｃｕｒｒｅｎｔ Ｐｒｏｔｏｃｏｌｓ ｉｎ Ｍｏｌｅｃｕｌａｒ Ｂｉｏｌｏｇｙ，Ｇｒｅｅｎｅ Ｐｕｂｌ．Ａｓｓｏｃ．＆Ｗｉｌｅｙ－Ｉｎｔｅｒｓｃｉｅｎｃｅｓ，Ｎｅｗ Ｙｏｒｋ ＮＹ、Ｉｎｎｉｓ，Ｍ．他（１９９０）ＰＣＲ Ｐｒｏｔｏｃｏｌｓ，Ａ Ｇｕｉｄｅ ｔｏ Ｍｅｔｈｏｄｓ ａｎｄ Ａｐｐｌｉｃａｔｉｏｎｓ，Ａｃａｄｅｍｉｃ Ｐｒｅｓｓ，Ｓａｎ Ｄｉｅｇｏ ＣＡなどを参照されたい。ＰＣＲプライマー対を既知の配列から得るには、例えば、そ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）を用い得る。

【００８２】

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および、最大５，０００までの大きめのポリヌクレオチドであって３２キロベースまでのインプットポリヌクレオチド配列から得た配列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、したがってゲノム全体の範囲でプライマーを設計するのに有用である。Ｐｒｉｍｅｒ３プライマー選択プログラム（Ｗｈｉｔｅｈｅａｄ Ｉｎｓｔｉｔｕｔｅ／ＭＩＴ Ｃｅｎｔｅｒ ｆｏｒ Ｇｅｎｏｍｅ Ｒｅｓｅａｒｃｈ（マサチューセッツ州ケンブリッジ）より入手可能）を用いれば、プライマー結合部位として避けたい配列を指定できる「非プライミングライブラリ（ｍｉｓｐｒｉｍｉｎｇ ｌｉｂｒａｒｙ）」を入力できる。Ｐｒｉｍｅｒ３は特に、マイクロアレイのためのオリゴヌクレオチドの選択に有用である（後二者のプライマー選択プログラムのソースコードは、それぞれの情報源から得てユーザー固有のニーズを満たすように修正し得る）。ＰｒｉｍｅｒＧｅｎプログラム（英国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ー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いず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。したがって、このプログラムは、独自のものであれ保存されたものであれ、オリゴヌクレオチドとポリヌクレオチド断片との同定に有用である。上記選択方法のいずれかによって同定したオリゴヌクレオチドおよびポリヌクレオチド断片は、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、ある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。

【００８３】

「組換え核酸」は、天然の配列ではなく、配列の、２つ以上の離れたセグメントを人工的に組み合わせた配列である。この人為的組合せはしばしば化学合成によって達成するが、より一般的には核酸の単離セグメントの人為的操作によって、例えばＳａｍｂｒｏｏｋの文献（前出）に記載されているような遺伝子工学的手法によって達成する。組換え核酸の語は、単に核酸の一部が付加、置換または欠失により改変された核酸も含む。しばしば組換え核酸には、プロモーター配列に機能的に連結した核酸配列が含まれる。このような組換え核酸は、例えば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成し得る。

【００８４】

あるいはこのような組換え核酸は、ウイルスベクターの一部と成すことができ、ベクターは例えばワクシニアウイルスに基づくものであり得る。そのようなベクターは哺乳類に接種され、その組換え核酸が発現されて、その哺乳類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。

【００８５】

「調節エレメント」は、或る遺伝子の非翻訳領域に通常は由来する核酸配列であり、エンハンサー、プロモーター、イントロンおよび５’および３’の非翻訳領域（ＵＴＲ）を含む。調節エレメントは、転写、翻訳またはＲＮＡ安定性を制御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。

【００８６】

「レポーター分子」は、核酸、アミノ酸または抗体の標識に用いる、化学的または生化学的な成分である。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、磁気粒子およびその他の当分野で既知の成分がある。

【００８７】

或るＤＮＡ配列に対する「ＲＮＡ等価物」は、基準となるＤＮＡ配列と同じ直鎖のヌクレオチド配列からなるが、生じる全ての窒素性塩基のチミンがウラシルに置換され、糖鎖のバックボーンがデオキシリボースではなくリボースからなる。

【００８８】

用語「サンプル」は、その最も広い意味で用いられている。ＡＴＲＳ、ＡＴＲＳをコードする核酸群、またはその断片群を含むと推定されるサンプルとしては、体液と、細胞や細胞から単離した染色体や細胞内小器官（オルガネラ）や膜からの抽出物と、細胞と、溶液中に存在するまたは基板に固定されたゲノムＤＮＡ、ＲＮＡ、ｃＤＮＡと、組織と、組織プリントなどがあり得る。

【００８９】

用語「特異結合」および「特異的に結合する」は、タンパク質若しくはペプチドの、アゴニスト、抗体、アンタゴニスト、小分子、若しくは任意の天然若しくは合成の結合組成物との間の相互作用を指す。この相互作用は、タンパク質の特定の構造（例えば抗原決定基すなわちエピトープ）であって結合分子が認識する構造の有無に依存する。例えば、或る抗体がエピトープ「Ａ」に対して特異的である場合、結合していない標識した「Ａ」と抗体とを含む或る反応の中に、エピトープＡを持つポリペプチドが、あるいは結合していない無標識の「Ａ」が存在すると、抗体と結合する標識Ａの量が減少する。

【００９０】

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

【００９１】

「置換」とは、１つ以上のアミノ酸残基またはヌクレオチドを、それぞれ別のアミノ酸残基またはヌクレオチドに置き換えることである。

【００９２】

用語「基板」は、任意の好適な固体あるいは半固体の支持物を指し、膜およびフィルター、チップ、スライド、ウエハ、ファイバー、磁気または非磁気ビーズ、ゲル、チューブ、プレート、ポリマー、微小粒子、毛細管が含まれる。基板は、ウェル、溝、ピン、チャネル、孔など、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。

【００９３】

「転写イメージ」または「発現プロフィール」は、所定条件下での所定時間における、或る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。

【００９４】

「形質転換」とは、外来性ＤＮＡが、或る受容細胞に導入されるプロセスを言う。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、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、リポフェクションおよび微粒子銃を用いる方法がある。「形質転換された細胞」には、導入されたＤＮＡが、自律的に複製するプラスミドとしてある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。更に、限られた期間に一過的に導入ＤＮＡ若しくは導入ＲＮＡを発現する細胞も含まれる。

【００９５】

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個以上の細胞が、ヒトの介入によって、例えば本技術分野で公知の形質転換技術によって導入された異種核酸を有するものである。細胞への核酸の導入は、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行う。これは、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ある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、古典的な交雑育種あるいは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。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、真菌および動植物がある。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公知であり、前出のＳａｍｂｒｏｏｋ ら （１９８９） 等の参考文献に記載がある。

【００９６】

特定の核酸配列の「変異体／変異配列」とは、核酸配列１本の或る長さ全体について、該特定核酸配列に対し少なくとも４０％の配列同一性を有する核酸配列として決定された配列である。決定には、デフォルトパラメータに設定した「ＢＬＡＳＴ ２ Ｓｅｑｕｅｎｃｅｓ」ツールＶｅｒｓｉｏｎ ２．０．９（１９９９年５月７日）を用いてｂｌａｓｔｎを実行する。このような核酸対は、所定の長さに対して、例えば少なくとも５０％、６０％、７０％、８０％、８５％、９０％、９１％、９２％、９３％、９４％、９５％、９６％、９７％、９８％、９９％またはそれ以上の配列同一性を示し得る。変異配列は、例えば、「対立遺伝子」変異配列（上述）または「スプライス」変異配列、「種」変異配列、「多型」変異配列と記述され得る。スプライス変異配列は参照分子との顕著な同一性を有

し得るが、ｍＲＮＡプロセッシング中のエキソン群の選択的スプライシングによって通常、より多数またはより少数のポリヌクレオチドを有することになる。対応するポリペプチドは、追加機能ドメイン群を有するか、あるいは参照分子には存在するドメイン群が欠落していることがある。種変異配列は、種によって異なるポリヌクレオチド配列である。結果的に生じるポリペプチドは通常、相互に有意のアミノ酸同一性を持つ。多型性変異配列は、或る種の個体間における、或る特定遺伝子のポリヌクレオチド配列内での変異である。多型変異配列にはまた、ポリヌクレオチド配列の１つのヌクレオチド塩基が異なる「１塩基多型性」（ＳＮＰ）も含み得る。ＳＮＰの存在は、例えば特定の個体群、病状または病状性向を示し得る。

【００９７】

特定のポリペプチド配列の「変異配列」とは、ポリペプチド配列１本の或る長さ全体について、該特定ポリペプチド配列に対し少なくとも４０％の配列同一性を有するポリペプチド配列として決定された配列である。決定には、デフォルトパラメータに設定した「ＢＬＡＳＴ２Ｓｅｑｕｅｎｃｅｓ」ツールＶｅｒｓｉｏｎ２．０．９（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、そのポリペプチドの一方の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。

【００９８】

（発明）

本発明は、新規のヒトのアミノアシルｔＲＮＡ 合成酵素（ＡＴＲＳ） 及びＡＴＲＳをコードするポリヌクレオチドの発見に基づき、また、これらの組成物を利用した細胞増殖異常、自己免疫／炎症性の疾患の診断、治療、及び予防の発見にも基づく。

【００９９】

表１は、本発明の完全長ポリヌクレオチド配列およびポリペプチド配列の命名の概略である。各ポリヌクレオチドおよ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）に相関する。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。

【０１００】

表２は、ＧｅｎＢａｎｋタンパク質（ｇｅｎｐｅｐｔ）データベースに対するＢＬＡＳＴ分析で同定した、本発明のポリペプチド群に相同な配列群を示す。列１および列２はそれぞれ、本発明の各ポリペプチドに対するポリペプチド配列識別番号（ポリペプチド ＳＥＱ ＩＤ ＮＯ：）と、それに対応するＩｎｃｙｔｅポリペプチド配列番号（ＩｎｃｙｔｅポリペプチドＩＤ）を示す。列３は、最も近いＧｅｎＢａｎｋ相同体のＧｅｎＢａｎｋ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１つ以上との間の一致に関する確率スコアを示す。列５は、該当箇所には適当な引用を示すとともにＧｅｎＢａｎｋ相同体１つ以上の注釈（ａｎｎｏｔａｔｉｏｎ）を示し、これらはすべて本明細書では参考文献に含まれる。

【０１０１】

表３は、本発明のポリペプチドの多様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、該当箇所には更に、分析方法に利用した検索可能なデータベースを示す。

【０１０２】

表２及び表３は共に、本発明のポリペプチドの特性を要約しており、それら特性が、請求の範囲に記載されたポリペプチドがアミノアシルｔＲＮＡ合成酵素であることを確立している。例えば、ＳＥＱ ＩＤ ＮＯ：１はマウスのバリルｔＲＮＡ合成酵素（ＧｅｎＢａｎｋ ＩＤ ｇ４５９０３２８）との５０％の同一性を有することがＢａｓｉｃ ＬｏｃａｌＡｌｉｇｎｍｅｎｔ Ｓｅａｒｃｈ Ｔｏｏｌ （ＢＬＡＳＴ）によって判定された（表２参照）。ＢＬＡＳＴ確率スコアは３．５ｅ－２７１であり、これは観測されたポリペプチド配列アラインメントが偶然に得られる確率を示している。ＳＥＱ ＩＤ ＮＯ：１はまた、ｔＲＮＡ 合成酵素クラスＩ （Ｉ、Ｌ、Ｍ、およびＶ）ドメインを持つ。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群の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判定された （表３参照）。ＢＬＩＭＰＳ、ＭＯＴＩＦＳ、及びＰＲＯＦＩＬＥＳＣＡＮ解析よりのデータは、ＳＥＱ ＩＤ ＮＯ：１がバリルｔＲＮＡ合成酵素である、さらに実証的な証拠を提供する。ＳＥＱ ＩＤ ＮＯ：１の解析用のアルゴリズム及びパラメータを表７に記述した。

【０１０３】

表４に示すように、本発明の完全長ポリヌクレオチド配列は、ｃＤＮＡ配列またはゲノムＤＮＡ由来のコード（エキソン）配列を用いて、あるいはこれら２種類の配列を任意に組み合わせて構築した。列１は本発明の各ポリヌクレオチドに対するポリヌクレオチド配列識別番号（ポリヌクレオチドＳＥＱ ＩＤ ＮＯ）および対応するＩｎｃｙｔｅポリヌクレオチドコンセンサス配列番号（Ｉｎｃｙｔｅ ＩＤ）、および塩基対の各ポリヌクレオチド配列の長さを示している。列２は、本発明の完全長ポリヌクレオチド配列を構築するのに用いたｃＤＮＡ配列および／またはゲノム配列の、また、例えばＳＥＱ ＩＤ ＮＯ：２を同定するため、或いはＳＥＱ ＩＤ ＮＯ：２と、関連するポリヌクレオチド配列群とを区別するためのハイブリダイゼーション技術または増幅技術に有用なポリヌクレオチド配列の断片の、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。

【０１０４】

表４の列２に記載したポリヌクレオチド断片は、特に例えば組織特異的ｃＤＮＡライブラリ群やプールしたｃＤＮＡライブラリ群に由来するＩｎｃｙｔｅ ｃＤＮＡ群を指す場合もある。或いは列２に記載したポリヌクレオチド断片は、完全長ポリヌクレオチド配列の構築（アセンブリ）に寄与したＧｅｎＢａｎｋ ｃＤＮＡまたはＥＳＴを指す場合もある。更に、列２に記載したポリヌクレオチド断片は、ＥＮＳＥＭＢＬ（Ｔｈｅ Ｓａｎｇｅｒ Ｃｅｎｔｒｅ、英国ケンブリッジ）データベースに由来する配列を同定する場合もある（すなわち「ＥＮＳＴ」の命名を含む配列）。あるいは、列２に記載したポリヌクレオチド断片は、ＮＣＢＩ ＲｅｆＳｅｑ Ｎｕｃｌｅｏｔｉｄｅ Ｓｅｑｕｅｎｃｅ Ｒｅｃｏｒｄｓ データベースに由来する場合もあり（すなわち「ＮＭ」または「ＮＴ」の命名を含む配列）、またＮＣＢＩ ＲｅｆＳｅｑ Ｐｒｏｔｅｉｎ Ｓｅｑｕｅｎｃｅ Ｒｅｃｏｒｄｓに由来する場合もある（すなわち「ＮＰ」の命名を含む配列）。または列２に記したポリヌクレオチド断片は、「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およびＧｅｎｓｃａｎ予測エキソン群の両方からなる集合を指す場合がある。例えば、ポリヌクレオチド配列の内、ＦＬ＿ＸＸＸＸＸ＿$Ｎ_1$＿$Ｎ_2$＿ＹＹＹＹＹ＿$Ｎ_3$＿$Ｎ_4$ として同定される配列は「スティッチされた」配列であり、その内、ＸＸＸＸＸＸは該アルゴリズムが適用される配列のクラスターの識別番号であり、ＹＹＹＹＹは該アルゴリズムが作成する予測の数であり、$Ｎ_{1,2,3...}$がある場合には、解析中に手動で編集された特定のエキソン群を表す（実施例５参照）。または、列２のポリヌクレオチド断片は「エキソンストレッチング（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場合もある。例えば、ポリヌクレオチド配列の内、ＦＬＸＸＸＸＸＸ＿ｇＡＡＡＡＡ＿ｇＢＢＢＢ

Ｂ＿１＿Ｎとして同定される配列は「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、Ｎ は特定のエキソンを指す（実施例５を参照）。或るＲｅｆＳｅｑ配列が「エキソンストレッチング」アルゴリズムのためのタンパク質相同体として使用された場合では、「ＮＭ」、「ＮＰ」、または「ＮＴ」によって表されるＲｅｆＳｅｑ識別子が、ＧｅｎＢａｎｋ識別子（すなわち、ｇＢＢＢＢＢ）の代わりに使用され得る。

【０１０５】

あるいは、接頭コードは、手動で編集された構成配列、ゲノムＤＮＡ配列から予測された構成配列、または組み合わされた配列解析方法に由来する構成配列を同定する。次の表は、構成配列の接頭コードと、接頭コードに対応する配列分析方法の例を列記する（実施例４と５を参照）。

接頭コード	解析タイプやプログラムの例
GNN, GFG, ENST	例えば、GENSCAN (Stanford University, CA, USA) または FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK) を用いた、ゲノム配列群からのエキソン予測
GBI	手動で編集された、ゲノム配列群の解析
FL	スティッチまたはストレッチゲノム配列 (実施例5参照)
INCY	ゲノムへのEST配列群のマッピングからの、完全長転写物とエキソンとの予想。エキソン群と生じる転写物とを予想するために、ゲノム位置とEST構成のデータが組み合わされる。

【０１０６】

場合によっては、最終コンセンサスポリヌクレオチド配列を確認するために、表４に示すような配列カバレッジと重複するＩｎｃｙｔｅ ｃＤＮＡカバレッジが得られたが、該当するＩｎｃｙｔｅ ｃＤＮＡ識別番号は示さなかった。

【０１０７】

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ライブラリとはＩｎｃｙｔｅ ｃＤＮＡライブラリであり、これは、最も頻繁にはＩｎｃｙｔｅ ｃＤＮＡ配列群によって代表されるが、これらは、上記のポリヌクレオチド配列を構築および確認するために用いられた。ｃＤＮＡライブラリを作製するために用いた組織およびベクターを表５に示し、表６で説明している。

【０１０８】

本発明にはまた、ＡＴＲＳ変異体をも含む。好適なＡＴＲＳ変異体は、ＡＴＲＳの機能的或いは構造的特徴の少なくとも１つを有し、かつ、該ＡＴＲＳアミノ酸配列に対して少なくとも約８０％のアミノ酸配列同一性、或いは少なくとも約９０％のアミノ酸配列同一性、更には少なくとも約９５％のアミノ酸配列同一性を有する変異体である。

【０１０９】

本発明はまた、ＡＴＲＳをコードするポリヌクレオチドを提供する。特定の実施態様において本発明は、ＡＴＲＳをコードする、ＳＥＱ ＩＤ ＮＯ：２の配列を持つポリヌクレオチド配列を提供する。配列表に示したように、ＳＥＱ ＩＤ ＮＯ：２のポリヌクレオチド配列には、窒素系塩基のチミンの出現がウラシルに置換され糖鎖のバックボーンがデオキシリボースではなくリボースからなる、等価ＲＮＡ配列をも含む。

【０１１０】

本発明にはまた、ＡＴＲＳをコードするポリヌクレオチド配列の変異配列を含む。詳細には、このような変異体ポリヌクレオチド配列は、ＡＴＲＳをコードする該ポリヌクレオチド配列との、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持つこととなる。本発明の或る特定の態様には、ＳＥＱ ＩＤ ＮＯ：２の配列を持つ或るポリヌクレオチド配列の変異体を含み、この変異体はＳＥＱ ＩＤ：２の核酸配列との、少なくとも約７０％、或いは少なくとも約８５％、または少なくとも約９５％ものポリヌクレオチド配列同一性を有する。上記の任意のポリヌクレオチド変異体は、ＡＴＲＳの機能的若しくは構造的特徴を少なくとも１つ有する或るアミノ酸配列をコードし得る。

【０１１１】

更に別の例では、本発明の或るポリヌクレオチド変異配列は、ＡＴＲＳをコードするポリヌクレオチド配列のスプライス変異配列である。或るスプライス変異配列は、ＡＴＲＳをコードするポリヌクレオチド配列との顕著な配列同一性を持つ部分複数を有し得るが、ｍＲＮＡプロセッシング中のエキソン群の選択的スプライシングによって生ずる、配列の数ブロックの付加または欠失により、通常、より多数またはより少数のポリヌクレオチドを有することになる。或るスプライス変異配列には、約７０％未満、または約６０％未満、あるいは約５０％未満のポリヌクレオチド配列同一性が、ＡＴＲＳをコードするポリヌクレオチド配列との間で全長に渡って見られるが、このスプライス変異配列のいくつかの部分には、ＤＭＥをコードするポリヌクレオチド配列の各部との、少なくとも約７０％、あるいは少なくとも約８５％、または少なくとも約９５％、なおまたは１００％の、ポリヌクレオチド配列同一性を有することとなる。上記したスプライス変異配列は何れも、ＡＴＲＳの機能的或いは構造的特徴の少なくとも１つを有する或るアミノ酸配列をコードし得る。

【０１１２】

遺伝暗号の縮重により、ＡＴＲＳをコードする種々のポリヌクレオチド配列が作られ、中には、既知のいかなる天然遺伝子のポリヌクレオチド配列群とも最小の類似性しか有しない配列もあることは、当業者には理解されよう。したがって本発明には、可能コドン選択に基づく組合せの選択によって産出し得るあらゆる可能なポリヌクレオチド配列のバリエーションを網羅し得る。これらの組み合わせは、天然ＡＴＲＳ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成される。全てのそのような変異が明確に開示されていると考慮されたい。

【０１１３】
ＡＴＲＳとその変異配列とをコードするヌクレオチド配列は一般に、好適に選択されたストリンジェンシー条件下で天然ＡＴＲＳのヌクレオチド配列とハイブリダイズ可能であるが、非天然コドン群を含めるなどの実質的に異なるコドン使用を有する、ＡＴＲＳ或いはその誘導体をコードするヌクレオチド配列を作り出すことは、有益であり得る。宿主が特定コドンを利用する頻度に基づいて、特定の真核宿主また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し得る。コードされるアミノ酸配列を改変せずに、ＡＴＲＳ及びその誘導体をコードするヌクレオチド配列を実質的に改変する別の理由には、天然配列から作られる転写物より例えば長い半減期など好ましい特性を備えるＲＮＡ転写物を作ることもある。

【０１１４】
本発明にはまた、ＡＴＲＳ及びその誘導体群をコードする、ＤＮＡ配列またはそれらの断片群を、完全に合成化学によって作り出す過程をも含む。作製後、当分野で周知の試薬を用いて、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ＡＴＲＳまたはその任意の断片をコードする或る配列に突然変異を誘導し得る。

【０１１５】
更に本発明には、種々のストリンジェンシー条件下で、請求項に記載されたポリヌクレオチド配列に、特にＳＥＱ ＩＤ ＮＯ：２及びそれらの断片にハイブリダイズ可能なポリヌクレオチド配列群が含まれる（例えば、Ｗａｈｌ，Ｇ．Ｍ．およびＳ．Ｌ．Ｂｅｒｇｅｒ（１９８７）Ｍｅｔｈｏｄｓ Ｅｎｚｙｍｏｌ．１５２：３９９－４０７；Ｋｉｍｍｅｌ．Ａ．Ｒ．（１９８７）Ｍｅｔｈｏｄｓ Ｅｎｚｙｍｏｌ．１５２：５０７－５１１を参照）。アニーリングおよ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」に記載した。

【０１１６】
ＤＮＡシークエンシングの方法は当分野では公知であり、本発明のいずれの実施例も、ＤＮＡシークエンシング方法を用い得る。ＤＮＡシークエンシング方法には酵素を用い得る。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）を用い得る。あるいは、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とを併用し得る。好適には、ＭＩＣＲＯＬＡＢ２２００液体転移システム（Ｈａｍｉｌｔｏｎ，Ｒｅｎｏ，ＮＶ）、ＰＴＣ２００サーマルサイクラー（ＭＪ Ｒｅｓｅａｒｃｈ，Ｗａｔｅｒｔｏｗｎ ＭＡ）およびＡＢＩ ＣＡＴＡＬＹＳＴ ８００サーマルサイクラー（Ａｐｐｌｉｅｄ Ｂｉｏｓｙｓｔｅｍｓ）などの装置を用いて配列の準備を自動化する。次に、ＡＢＩ ３７３ あるいは 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）または当分野で公知の他のシステムを用いてシークエンシングを行う。結果として得た配列を、当分野で周知の種々のアルゴリズムを用いて分析する（例えば、Ａｕｓｕｂｅｌ，Ｆ．Ｍ．（１９９７）Ｓｈｏｒｔ Ｐｒｏｔｏｃｏｌｓ ｉｎ Ｍｏｌｅｃｕｌａｒ Ｂｉｏｌｏｇｙ，Ｊｏｈｎ Ｗｉｌｅｙ ＆ Ｓｏｎｓ，Ｎｅｗ Ｙｏｒｋ ＮＹ，７．７ユニット、Ｍｅｙｅｒｓ，Ｒ．Ａ．（１９９５）Ｍｏｌｅｃｕｌａｒ Ｂｉｏｌｏｇｙ ａｎｄ Ｂｉｏｔｅｃｈｎｏｌｏｇｙ，Ｗｉｌｅｙ ＶＣＨ，Ｎｅｗ Ｙｏｒｋ ＮＹ，８５６－８５３ページを参照）。

【０１１７】
当分野で周知の、ＰＣＲ法をベースにした種々の方法と、部分的ヌクレオチド配列とを利用して、ＡＴＲＳ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、

上流にある配列を検出し得る。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およ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からの未知の配列を増幅する方法である（例えば、Ｓａｒｋａｒ，Ｇ．（１９９３）ＰＣＲ Ｍｅｔｈｏｄｓ Ａｐｐｌｉｃ．２：３１８－３２２を参照）。別の方法に逆ＰＣＲ法があり、これは多岐の方向に伸長するプライマー群を用いて、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、或る既知のゲノム遺伝子座およびその周辺の配列群からなる制限酵素断片群から得る（例えば、Ｔｒｉｇｌｉａ，Ｔ．他（１９８８）Ｎｕｃｌｅｉｃ Ａｃｉｄｓ Ｒｅｓ．１６：８１８６を参照）。第３の方法としてキャプチャＰＣＲ法があり、これにはヒトおよび酵母人工染色体ＤＮＡの既知の配列群に隣接するＤＮＡ断片群をＰＣＲ増幅する方法を含む（例えばＬａｇｅｒｓｔｒｏｍ，Ｍ．他（１９９１）ＰＣＲ Ｍｅｔｈｏｄｓ Ａｐｐｌｉｃ １：１１１－１１９を参照）。この方法では、ＰＣＲを行う前に、複数の制限酵素の消化およびライゲーション反応を用い、未知の配列領域内に組換え二本鎖配列を挿入し得る。未知の配列群を検索するために用い得る他の複数の方法も当分野で公知である（例えばＰａｒｋｅｒ，Ｊ．Ｄ．他（１９９１）Ｎｕｃｌｅｉｃ Ａｃｉｄｓ Ｒｅｓ．１９：３０５５－３０６０を参照）。更に、ＰＣＲ、ネステッドプライマー類、およびＰＲＯＭＯＴＥＲＦＩＮＤＥＲライブラリ類（Ｃｌｏｎｔｅｃｈ，Ｐａｌｏ Ａｌｔｏ ＣＡ）を用いてゲノムＤＮＡをウォーキングし得る。この手順は、ライブラリ類をスクリーニングする必要がなく、イントロン／エキソン接合部の発見に有用である。全てのＰＣＲベースの方法では、市販ソフトウェア、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％以上で、温度約６８℃～７２℃で鋳型に対してアニーリングするように、プライマー群を設計し得る。

【０１１８】

完全長ｃＤＮＡ群をスクリーニングする際は、より大きなｃＤＮＡ群を含むようにサイズ選択されたライブラリ群を用いるのが好ましい。更に、ランダムプライマーのライブラリ群は、しばしば遺伝子群の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。ゲノムライブラリ群は、５’非転写調節領域への、配列の伸長に有用であろう。

【０１１９】

市販のキャピラリー電気泳動システム群を用いて、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し得る。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、発光された波長の検出に利用するＣＣＤカメラとを有し得る。出力／光の強度は、適切なソフトウェア（Ａｐｐｌｉｅｄ Ｂｉｏｓｙｓｔｅｍｓ社のＧＥＮＯＴＹＰＥＲ、ＳＥＱＵＥＮＣＥ ＮＡＶＩＧＡＴＯＲなど）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およ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。

【０１２０】

本発明の別の実施態様では、ＡＴＲＳをコードするポリヌクレオチド配列群またはその断片群を組換えＤＮＡ分子群にクローニングして、適切な宿主細胞内にＡＴＲＳ、その断片または機能的等価物を発現させ得る。遺伝暗号固有の縮重により、実質的に同じ或いは機能的に等価の或るアミノ酸配列をコードする別のＤＮＡ配列群を作ることができ、これらの配列をＡＴＲＳの発現に利用し得る。

【０１２１】

ＡＴＲＳをコードする配列を種々の目的で改変するために、当分野で一般的に既知の複数の方法を用いて、本発明のヌクレオチド配列を組換え得る。これらの目的には、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。遺伝子断片と合成オリゴヌクレオチドとの、ランダムなフラグメンテーションおよび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得る。例えば、オリゴヌクレオチドを介した部位特異的変異誘導を利用して、新規な制限部位の作製、グリコシル化パターンの改変、コドン優先の変更、スプライス変異配列の生成などを起こす突然変異を導入し得る。

【０１２２】

本発明のヌクレオチドは、ＭＯＬＥＣＵＬＡＲＢＲＥＥＤＩＮＧ（Ｍａｘｙｇｅｎ Ｉｎｃ．，Ｓａｎｔａ Ｃｌａｒａ ＣＡ；米国特許第５，８３７，４５８号；Ｃｈａｎｇ，Ｃ．－Ｃ．他（１９９９）Ｎａｔ．Ｂｉｏｔｅｃｈｎｏｌ．１７：７９３－７９７；Ｃｈｒｉｓｔｉａｎｓ，Ｆ．Ｃ．他（１９９９）Ｎａｔ．Ｂｉｏｔｅｃｈｎｏｌ．１７：２５９－２６４；Ｃｒａｍｅｒｉ，Ａ．他（１９９６）Ｎａｔ．Ｂｉｏｔｅｃｈｎｏｌ．１４：３１５－３１９）などのＤＮＡシャッフリング技術の対象となり得るもので、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、ＡＴＲＳの生物学的特性を改変或いは改良し得る。ＤＮＡシャッフリングは、ＰＣＲを介する遺伝子断片組換えを用いて遺伝子変異配列のライブラリを作製するプロセスである。ライブラリはその後、所望の特性を持つ遺伝子変異配列群を同定する、選択またはスクリーニングの手順を経る。続いて、これら好適な変異配列をプールし、更に反復してＤＮＡシャッフリングおよび選択／スクリーニングを行い得る。かくして、「人工的な」育種および急速な分子の進化によって、多様な遺伝子が作られる。例えば、ランダムポイント突然変異を持つ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し得る。あるいは、所定の遺伝子の断片群を同種または異種のいずれかから得た同一遺伝子ファミリーの相同遺伝子の断片と組み換え、それによって天然に存在する複数の遺伝子の遺伝多様性を、方向付けした制御可能な方法で最大化させることができる。

【０１２３】

別の実施態様では、ＡＴＲＳ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され得る（例えば、Ｃａｒｕｔｈｅｒｓ．Ｍ．Ｈ．他（１９８０）（１９８０）Ｎｕｃｌ．Ａｃｉｄｓ Ｒｅｓ．Ｓｙｍｐ．Ｓｅｒ ７：２１５－２２３；及びＨｏｒｎ，Ｔ．他（１９８０）Ｎｕｃｌ．Ａｃｉｄｓ Ｒｅｓ．Ｓｙｍｐ．Ｓｅｒ．７：２２５－２３２を参照）。別法として、化学的方法を用いてＡＴＲＳ自体またはその断片を合成し得る。例えば、種々の液相または固相技術を用いてペプチド合成を行い得る（例えばＣｒｅｉｇｈｔｏｎ，Ｔ．（１９８４）Ｐｒｏｔｅｉｎｓ，Ｓｔｒｕｃｔｕｒｅｓ ａｎｄ Ｍｏｌｅｃｕｌａｒ Ｐｒｏｐｅｒｔｉｅｓ，ＷＨ Ｆｒｅｅｍａｎ，Ｎｅｗ Ｙｏｒｋ ＮＹ，５５－６０ページ、Ｒｏｂｅｒｇｅ，Ｊ．Ｙ．他（１９９５）Ｓｃｉｅｎｃｅ ２６９：２０２－２０４を参照）。自動合成はＡＢＩ ４３１Ａペプチドシンセサイザ（Ａｐｐｌｉｅｄ Ｂｉｏｓｙｓｔｅｍｓ）を用いて達成し得る。更に、ＡＴＲＳのアミノ酸配列または任意のその一部を、直接合成の際に改変することにより、及び／または他のタンパク質からの配列群または任意のその一部と組み合わせることにより、或る天然ポリペプチドの配列を有するポリペプチドまたは変異体ポリペプチドが作製され得る。

【０１２４】

このペプチドは、分離用高速液体クロマトグラフィーを用いて実質的に精製し得る（例えばＣｈｉｅｚ，Ｒ．Ｍ．およびＦ．Ｚ．Ｒｅｇｎｉｅｒ（１９９０）Ｍｅｔｈｏｄｓ Ｅｎｚｙｍｏｌ．１８２：３９２－４２１を参照）。この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できる（例えば前出のＣｒｅｉｇｈｔｏｎ，２８－５３ページを参照）。

【０１２５】

生物学的に活性なＡＴＲＳを発現させるために、ＡＴＲＳまたはその誘導体群をコードするヌクレオチド配列群を適切な発現ベクターに挿入し得る。適切な発現ベクターとは、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制御に必要なエレメント群を持

つベクターである。これらのエレメントの例には、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、ベクターに及びＡＴＲＳをコードするポリヌクレオチド配列群における調節配列が含まれる。必要なエレメント群は、特長および特異性が様々である。特定の開始シグナル類によって、ＡＴＲＳをコードする配列群の、より効果的な翻訳を達成し得る。開始シグナルの例には、ＡＴＧ開始コドンと、コザック配列など近傍の配列とが含まれる。ＡＴＲＳをコードする配列群及びその開始コドン、上流の調節配列群が好適な発現ベクターに挿入された場合は、更なる転写制御シグナルや翻訳制御シグナルは必要ないこともある。しかしながら、コーディング配列あるいはその断片のみが挿入された場合は、インフレームＡＴＧ開始コドンなど外来性の翻訳制御シグナルが発現ベクターに含まれるようにすべきである。外来性の翻訳エレメント群および開始コドン群は、様々な天然物および合成物を起源とし得る。用いられる特定の宿主細胞系に好適なエンハンサーを含めることで発現の効率を高めることが可能である（例えばＳｃｈａｒｆ，Ｄ．他（１９９４）Ｒｅｓｕｌｔｓ Ｐｒｏｂｌ．Ｃｅｌｌ Ｄｉｆｆｅｒ．２０：１２５－１６２を参照）。

【０１２６】

当業者に周知の方法を用いて、ＡＴＲＳをコードする配列群と、好適な転写及び翻訳制御エレメントとを持つ発現ベクター群を作製し得る。これらの方法としては、ｉｎ ｖｉｔｒｏ組換えＤＮＡ技術、合成技術、およびｉｎ ｖｉｖｏ遺伝子組換え技術がある（例えば、Ｓａｍｂｒｏｏｋ，Ｊ．他．（１９８９）Ｍｏｌｅｃｕｌａｒ Ｃｌｏｎｉｎｇ，Ａ Ｌａｂｏｒａｔｏｒｙ Ｍａｎｕａｌ，Ｃｏｌｄ Ｓｐｒｉｎｇ Ｈａｒｂｏｒ Ｐｒｅｓｓ，Ｐｌａｉｎｖｉｅｗ ＮＹ，４章および８章，および１６－１７章；およびＡｕｓｕｂｅｌ，Ｆ．Ｍ．他．（１９９５）Ｃｕｒｒｅｎｔ Ｐｒｏｔｏｃｏｌｓ ｉｎ Ｍｏｌｅｃｕｌａｒ Ｂｉｏｌｏｇｙ，Ｊｏｈｎ Ｗｉｌｅｙ ＆ Ｓｏｎｓ，Ｎｅｗ Ｙｏｒｋ ＮＹ，９章、１３章および１６章を参照）。

【０１２７】

種々の発現ベクター／宿主系を利用して、ＡＴＲＳをコードする配列の保持及び発現ができる。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など微生物や、ウイルス発現ベクター（例えばバキュロウイルス）に感染した昆虫細胞系や、ウイルス発現ベクター（例えばカリフラワーモザイクウイルス、ＣａＭＶまたはタバコモザイクウイルス、ＴＭＶ）または細菌発現ベクター（例えばＴｉプラスミドまたはｐＢＲ３２２プラスミド）で形質転換させた植物細胞系、動物細胞系がある（例えば前出のＳａｍｂｒｏｏｋ、前出のＡｕｓｕｂｅｌ、Ｖａｎ Ｈｅｅｋｅ，Ｇ．およびＳ．Ｍ．Ｓｃｈｕｓｔｅｒ（１９８９）Ｊ．Ｂｉｏｌ．Ｃｈｅｍ．２６４：５５０３－５５０９、Ｅｎｇｅｌｈａｒｄ、Ｅ．Ｋ．他（１９９４）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ ９１：３２２４－３２２７、Ｓａｎｄｉｇ，Ｖ．他（１９９６）Ｈｕｍ．Ｇｅｎｅ Ｔｈｅｒ．７：１９３７－１９４５、Ｔａｋａｍａｔｓｕ，Ｎ．（１９８７）ＥＭＢＯ Ｊ．６：３０７－３１１、『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を参照）。レトロウイルス、アデノウイルス、ヘルペスウイルスまたはワクシニアウイルス由来の発現ベクター、または種々の細菌性プラスミド由来の発現ベクターを用いて、ヌクレオチド配列を標的器官、組織または細胞集団へ送達することができる（例えばＤｉ Ｎｉｃｏｌａ，Ｍ．他（１９９８）Ｃａｎｃｅｒ Ｇｅｎ．Ｔｈｅｒ．５（６）：３５０－３５６、Ｙｕ，Ｍ．他（１９９３）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ ９０（１３）：６３４０－６３４４、Ｂｕｌｌｅｒ，Ｒ．Ｍ．他（１９８５）Ｎａｔｕｒｅ ３１７（６０４０）：８１３－８１５；ＭｃＧｒ

egor, D. P. 他（1994）Mol. Immunol. 31（3）：219－226、Verma, I. M. and N. Somia（1997）Nature 389：239－242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。

【0128】

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を、ATRS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できる。例えば、ATRSをコードするポリヌクレオチド配列の慣例的なクローニング、サブクローニング、増殖には、PBLUESCRIPT（Stratagene, La Jolla CA）またはPSPORT1プラスミド（Life Technologies）など、多機能の大腸菌ベクターを用い得る。このベクターのマルチクローニング部位に、ATRS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、組換え分子を持つ形質転換された細菌の同定のための比色スクリーニング法が可能となる。更にこれらのベクターは、クローニングされた配列におけるin vitro転写、ジデオキシのシークエンシング、ヘルパーファージによる一本鎖のレスキュー、入れ子状態の欠失の生成にも有用であろう（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ATRSが必要な場合は、ATRS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持つベクターが使用できる。

【0129】

ATRSの産生には、酵母発現系を用い得る。α因子、アルコールオキシダーゼ、PGHプロモーターなど、構成型あるいは誘導型のプロモーターを持つ多数のベクターが、出芽酵母菌（Saccharomyces cerevisiae）またはピキア酵母（Pichia pastoris）内で使用可能である。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での保持かのどちらかを指示し、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ことを可能にする（例えば、Ausubel, 1995, 前出、Bitter, G. A. 他（1987）Methods Enzymol. 153：516－544、およびScorer. C. A. 他（1994）Bio／Technology 12：181－184を参照）。

【0130】

植物系もATRSの発現に用い得る。ATRS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、単独で或いはTMV（Takamatsu, N.（1987）EMBO J 6：307－311）由来のオメガリーダー配列と組み合わせて用いられるようなCaMV由来の35S及び19Sプロモーターによって駆動される。あるいは、RuBisCOの小サブユニットなど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得る（例えば、Coruzzi, G. 他.（1984）EMBO J. 3：1671－1680；Broglie, R. 他（1984）Science 224：838－843；およびWinter, J. 他（1991）Results Probl. Cell Differ. 17：85－105を参照）。これらの構成物は、直接DNA形質転換により、または病原体を媒介とする形質移入により、植物細胞内に導入し得る（例えば『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ページを参照）。

【0131】

哺乳類細胞においては、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、後期プロモーター及び3連リーダー配列からなるアデノウイルス転写／翻訳複合体に、ATRSをコードする配列をライゲーションし得る。アデノウイルスゲノムの可欠E1またはE3領域への挿入を用い、宿主細胞内でATRSを発現する感染ウイルスを得ることができる（例えばLogan, J. およびT. Sh

ｅｎｋ（１９８４）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ８１：３６５３６５９を参照）。更に、ラウス肉腫ウイルス（ＲＳＶ）エンハンサーなどの転写エンハンサーを用いて、哺乳動物宿主細胞における発現を増大させ得る。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。

【０１３２】

また、ヒト人工染色体（ＨＡＣ）類を用いて、或るプラスミドに含まれ得る断片やプラスミドから発現し得る断片より大きなＤＮＡの断片群を送達し得る。治療のために約６ｋｂ～１０ＭｂのＨＡＣが作製され、従来の送達方法（リポソーム、ポリカチオンアミノポリマー、またはベシクル）で送達されている（例えばＨａｒｒｉｎｇｔｏｎ．Ｊ．Ｊ．他（１９９７）Ｎａｔ Ｇｅｎｅｔ．１５：３４５－３５５を参照）。

【０１３３】

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、細胞株におけるＡＴＲＳの安定した発現が望ましい。例えば、ＡＴＲＳをコードする配列を細胞株に形質転換するために、発現ベクター類と、同じベクター上の或いは別のベクター上の選択可能マーカー遺伝子とを用い得る。用いる発現ベクターは、ウイルス起源の複製要素、および／または内因性の発現エレメント群を持ち得る。ベクターの導入後、選択培地に移す前に強化培地で約１～２日間、細胞を増殖させ得る。選択可能マーカーの目的は選択剤への抵抗性を与えることであり、選択可能マーカーの存在により、導入した配列をうまく発現するような細胞の成長およ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。

【０１３４】

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、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、抗生物質ある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および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（例えばＷｉｇｌｅｒ，Ｍ．他（１９８０）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ７７：３５６７－３５７０、Ｃｏｌｂｅｒｅ－Ｇａｒａｐｉｎ，Ｆ．他（１９８１）Ｊ．Ｍｏｌ．Ｂｉｏｌ．１５０：１－１４を参照）。この他の選択可能な遺伝子、例えば、代謝のための、細胞の必要条件を改変するｔｒｐＢおよびｈｉｓＤも、文献に記載がある（例えばＨａｒｔｍａｎ，Ｓ．Ｃ．およびＲ．Ｃ．Ｍｕｌｌｉｇａｎ（１９８８）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ８５：８０４７８０５１を参照）。可視マーカー類、例えばアントシアニンや、緑色蛍光タンパク質（ＧＦＰ；Ｃｌｏｎｔｅｃｈ）、βグルクロニダーゼおよびその基質であるβグルクロニド、またはルシフェラーゼおよびその基質であるルシフェリンを用い得る。これらのマーカーを用いて、トランスフォーマントを同定するだけでなく、特定のベクター系に起因する一過性あるいは安定したタンパク質発現を定量し得る（例えばＲｈｏｄｅｓ，Ｃ．Ａ．（１９９５）Ｍｅｔｈｏｄｓ Ｍｏｌ．Ｂｉｏｌ．５５：１２１１３１を参照）。

【０１３５】

マーカー遺伝子発現の有無によって目的の遺伝子の存在が示唆されても、その遺伝子の存在および発現の確認が必要な場合もある。例えば、ＡＴＲＳをコードする配列が或るマーカー遺伝子配列の中に挿入された場合、ＡＴＲＳをコードする配列を持つ形質転換された細胞群は、マーカー遺伝子機能の欠落により同定できる。または、単一プロモーターの制御下で、或るマーカー遺伝子を、ＡＴＲＳをコードする１配列とタンデムに配置できる。誘導または選択に応答したマーカー遺伝子の発現は通常、タンデム遺伝子の発現も示す。

【０１３６】
一般に、ＡＴＲＳをコードする核酸配列を持ち、ＡＴＲＳ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同定し得る。 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増幅や、核酸配列或いはタンパク質配列の検出及び／または数量化のための膜系、溶液ベース、或いはチップベースの技術を含むタンパク質生物学的試験法または免疫学的アッセイが含まれるが、これらに限定されるものではない。
【０１３７】
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、ＡＴＲＳの発現の検出及び計測のための免疫学的な方法は、当分野で既知である。このような技法の例には、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）、蛍光活性化細胞選別（ＦＡＣＳ、フローサイトメトリー）などがある。ＡＴＲＳ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群を用いた、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、競合結合試験を用いることもできる。これらのアッセイおよびこれ以外のアッセイは、当分野で公知である（例えばＨａｍｐｔｏｎ，Ｒ．他（１９９０）Ｓｅｒｏｌｏｇｉｃａｌ Ｍｅｔｈｏｄｓ，ａ Ｌａｂｏｒａｔｏｒｙ Ｍａｎｕａｌ，ＡＰＳ Ｐｒｅｓｓ．Ｓｔ．Ｐａｕｌ．ＭＮ，Ｓｅｃｔ．ＩＶ、Ｃｏｌｉｇａｎ，Ｊ．Ｅ．他（１９９７）Ｃｕｒｒｅｎｔ Ｐｒｏｔｏｃｏｌｓ ｉｎ Ｉｍｍｕｎｏｌｏｇｙ，Ｇｒｅｅｎｅ Ｐｕｂ．Ａｓｓｏｃｉａｔｅｓ ａｎｄ Ｗｉｌｅｙ－Ｉｎｔｅｒｓｃｉｅｎｃｅ，Ｎｅｗ Ｙｏｒｋ ＮＹ、Ｐｏｕｎｄ，Ｊ．Ｄ．（１９９８）Ｉｍｍｕｎｏｃｈｅｍｉｃａｌ Ｐｒｏｔｏｃｏｌｓ，Ｈｕｍａｎａ Ｐｒｅｓｓ，Ｔｏｔｏｗａ ＮＪを参照）。
【０１３８】
多岐にわたる標識方法および抱合方法が当業者に知られており、様々な核酸アッセイおよびアミノ酸アッセイにこれらの方法を用い得る。ＡＴＲＳ をコードするポリヌクレオチドに関連する配列を検出するための、標識されたハイブリダイゼーションプローブ或いはＰＣＲプローブを生成する方法には、オリゴ標識化、ニックトランスレーション、エンドラベリング（末端標識化）、または標識されたヌクレオチドを用いるＰＣＲ増幅が含まれる。別法として、ＡＴＲＳ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し得る。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などの好適なＲＮＡポリメラーゼおよ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など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あるいは標識には、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、発色剤などがある。
【０１３９】
ＡＴＲＳをコードするヌクレオチド配列で形質転換された宿主細胞は、細胞培地でのこのタンパク質の発現及び回収に好適な条件下で培養され得る。形質転換細胞から製造されたタンパク質が分泌されるか細胞内に留まるかは、使用される配列、ベクター、あるいはその両者に依存する。ＡＴＲＳをコードするポリヌクレオチドを持つ発現ベクターは、原核細胞膜または真核細胞膜を透過するＡＴＲＳの分泌を指示するシグナル配列群を持つように設計できることは、当業者には理解されよう。
【０１４０】
更に、宿主細胞株の選択は、挿入した配列の発現をモジュレートする能力、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

、脂質化およびアシル化がある。タンパク質の「プレプロ」または「プロ」形を切断する翻訳後のプロセシングを利用して、タンパク質の標的への誘導、折りたたみ、および／または活性を特定することが可能である。翻訳後の活性のための、特定の細胞装置および特徴的な機序を持つ、種々の宿主細胞（例えばＣＨＯ、ＨｅＬａ、ＭＤＣＫ、ＭＥＫ２９３、ＷＩ３８など）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および処理を確実にするように選択し得る。

【０１４１】

本発明の別の実施態様では、ＡＴＲＳをコードする天然核酸配列、修飾された核酸配列、または組換え核酸配列を、上記した任意の宿主系の融合タンパク質の翻訳を生じる異種配列にライゲーションさせ得る。例えば、市販の抗体によって認識できる異種部分を持つキメラＡＴＲＳタンパク質は、ＡＴＲＳ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および異種ペプチド部分も、市販されている親和性マトリックスを用いて融合タンパク質の精製を促進し得る。限定されるものではないがこのような部分には、グルタチオンＳトランスフェラーゼ（ＧＳＴ）、マルトース結合タンパク質（ＭＢＰ）、チオレドキシン（Ｔｒｘ）、カルモジュリン結合ペプチド（ＣＢＰ）、６－Ｈｉｓ、ＦＬＡＧ、ｃ－ｍｙｃ、赤血球凝集素（ＨＡ）がある。ＧＳＴは固定化グルタチオン上で、ＭＢＰはマルトース上で、Ｔｒｘはフェニルアルシンオキシド上で、ＣＢＰはカルモジュリン上で、そして６－Ｈｉｓは金属キレート樹脂上で、同族の融合タンパク質の精製を可能にする。ＦＬＡＧ、ｃ－ｍｙｃおよび赤血球凝集素（ＨＡ）は、これらのエピトープ標識を特異的に認識する市販されているモノクローナル抗体およびポリクローナル抗体を用いた、融合タンパク質の免疫親和性精製を可能にする。また、ＡＴＲＳをコードする配列と異種タンパク質配列との間に、タンパク質分解切断部位を融合タンパク質が持つように遺伝子操作すると、ＡＴＲＳが、精製後に異種部分から切断され得る。融合タンパク質の発現および精製方法は、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および精製を促進することもできる。

【０１４２】

本発明の別の実施態様では、ＴＮＴウサギ網状赤血球可溶化液またはコムギ胚芽抽出系（Ｐｒｏｍｅｇａ）を用いてｉｎ ｖｉｔｒｏで、放射能標識したＡＴＲＳの合成が可能である。これらの系は、Ｔ７、Ｔ３またはＳＰ６プロモーターと機能的に連結したタンパク質コード配列の、転写と翻訳とを結合する。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。

【０１４３】

本発明のＡＴＲＳまたはその断片を用いて、ＡＴＲＳに特異結合する化合物をスクリーニングできる。少なくとも１つまたは複数の試験化合物を用いて、ＡＴＲＳへの特異的な結合をスクリーニングし得る。試験化合物としては、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。

【０１４４】

一実施態様では、このように同定された化合物は、ＡＴＲＳの天然リガンドに密接に関連し、例えばリガンドやその断片であり、または天然基質や、構造的または機能的な擬態物質（ｍｉｍｅｔｉｃ）あるいは自然結合パートナーである（例えばＣｏｌｉｇａｎ，Ｊ．Ｅ．他（１９９１）Ｃｕｒｒｅｎｔ Ｐｒｏｔｏｃｏｌｓ ｉｎ Ｉｍｍｕｎｏｌｏｇｙ １（２）の５章を参照）。同様に、化合物は、ＡＴＲＳ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。一実施態様では、このような化合物に対するスクリーニングには、分泌タンパク質として或いは細胞膜上でＡＴＲＳ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、哺乳類、酵母、ショウジョウバエ、または大腸菌からの細胞が含まれる。ＡＴＲＳを発現する細胞群またはＡＴＲＳを含有する細胞膜分画を次に試験化合物と接触させて、結合や刺激、若しくは

ＡＴＲＳまたは化合物の何れかの活性の阻害を分析する。

【０１４５】

或るアッセイは、単に試験化合物をポリペプチドに実験的に結合させ、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その結合を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ＡＴＲＳと結合させるステップと、ＡＴＲＳ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および測定を行うことができる。更にこのアッセイは、無細胞再構成標本、化学ライブラリ、または、天然産物の混合物を用いて実施でき、試験化合物（群）は、溶液中で遊離させるか固体支持体に固定する。

【０１４６】

本発明のＡＴＲＳまたはその断片を用いて、ＡＴＲＳの活性を調節する化合物をスクリーニングできる。このような化合物の例には、アゴニスト、アンタゴニスト、部分的アゴニストまたは逆アゴニストなどが含まれる。一実施態様では、ＡＴＲＳが少なくとも１つの試験化合物と混合される、ＡＴＲＳの活性が許容される条件下でアッセイを実施し、試験化合物の存在下でのＡＴＲＳの活性を、試験化合物不在下でのＡＴＲＳの活性と比較する。試験化合物の存在下でのＡＴＲＳの活性の変化は、ＡＴＲＳの活性を調節する化合物の存在を標示する。別法では、試験化合物を、ＡＴＲＳの活性に適した条件下で、ＡＴＲＳを有するｉｎ ｖｉｔｒｏ系すなわち無細胞系と混合してアッセイを実施する。これらアッセイの何れにおいても、ＡＴＲＳの活性を調節する試験化合物は間接的に調節する場合があり、その際は試験化合物と直接接触する必要がない。少なくとも１つ、または複数の試験化合物をスクリーニングすることができる。

【０１４７】

別の実施態様では、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、ＡＴＲＳをコードするポリヌクレオチドまたはその哺乳類相同体を「ノックアウト」し得る。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産生に有用である（米国特許第５，１７５，３８３号および第５，７６７，３３７号などを参照）。例えば１２９／ＳｖＪ細胞系などのマウスＥＳ細胞は初期のマウス胚に由来し、培地で増殖させることができる。このＥＳ細胞は、ネオマイシンホスホトランスフェラーゼ遺伝子（ｎｅｏ：Ｃａｐｅｃｃｈｉ，Ｍ．Ｒ．（１９８９）Ｓｃｉｅｎｃｅ ２４４：１２８８－１２９２）などのマーカー遺伝子で破壊した、目的の遺伝子を含むベクターで形質転換する。このベクターは、相同組換えにより、宿主ゲノムの対応する領域に組み込まれる。別法では、Ｃｒｅ－ｌｏｘＰ系を用いて相同組換えを行い、組織特異的または発生段階特異的に目的遺伝子をノックアウトする（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マウス株などから採取したマウス細胞胚盤胞に微量注入する。これらの胚盤胞を偽妊娠メスに外科的に導入し、得られるキメラ子孫の遺伝形質を特定し、これらを交配させてヘテロ接合性系またはホモ接合性系を作製する。このようにして産生した遺伝子組換え動物は、潜在的治療薬や毒性薬剤で検査することができる。

【０１４８】

ＡＴＲＳ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細胞において操作し得る。ヒトＥＳ細胞は、内胚葉、中胚葉および外胚葉の細胞タイプを含む、少なくとも８つの別々の細胞系統に分化する可能性を有する。これらの細胞系統は、例えば神経細胞、造血系統および心筋細胞に分化する（Ｔｈｏｍｓｏｎ，Ｊ．Ａ．他（１９９８）Ｓｃｉｅｎｃｅ ２８２：１１４５－１１４７）。

【０１４９】

ＡＴＲＳをコードするポリヌクレオチドを用いて、ヒト疾患をモデル化した「ノックイン」ヒト化動物（ブタ）または遺伝子組換え動物（マウスまたはラット）を作製し得る。ノ

ックイン技術を用いて、ＡＴＲＳ をコードするポリヌクレオチドの或る領域を動物ＥＳ細胞に注入し、注入した配列を動物細胞ゲノムに組み込ませる。形質転換細胞を胞胚に注入し、胞胚を上記のように移植する。遺伝子組換え子孫または近交系について試験し、潜在的医薬品を用いて処理し、ヒトの疾患の治療に関する情報を得る。別法では、例えばＡＴＲＳを乳汁内に分泌するなどＡＴＲＳ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を参照）。

【０１５０】

（治療）

ＡＴＲＳの複数の領域とアミノアシルｔＲＮＡ合成酵素類との間には、例えば配列及びモチーフの文脈における化学的及び構造的類似性が存在する。また、ＡＴＲＳを発現する組織の数例は、表６を見られたい。このように、ＡＴＲＳは、細胞増殖異常および、自己免疫／炎症疾患において、或る役割を果たすと考えられる。ＡＴＲＳ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、ＡＴＲＳの発現または活性を低下させることが望ましい。また、ＡＴＲＳ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、ＡＴＲＳの発現または活性を増大させることが望ましい。

【０１５１】

従って、一実施態様において、ＡＴＲＳの発現または活性の低下に関連した疾患の治療または予防のために、被験者にＡＴＲＳまたはその断片や誘導体を投与し得る。限定するものではないが、このような疾患には、細胞増殖異常として、日光角化症、動脈硬化、アテローム性硬化、滑液包炎、硬変、肝炎、混合性結合組織病（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、また腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫及び奇形癌など癌、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、子宮の癌があり、自己免疫／炎症性疾患には、後天性免疫不全症候群（ＡＩＤＳ）、アジソン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血症、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫性多腺性内分泌カンジダ症外胚葉ジストロフィ（ＡＰＥＣＥＤ）、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー皮膚炎、皮膚筋炎、真性糖尿病、肺気腫、リンパ球傷害因子性偶発性リンパ球減少症、新生児溶血性疾患（胎児赤芽球症）、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性腸症候群、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、慢性関節リウマチ、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、ぶどう膜炎、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、外傷が含まれる。

【０１５２】

別の実施態様では、限定するものではないが上に列記した疾患を含むＡＴＲＳ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、ＡＴＲＳまたはその断片や誘導体を発現し得るベクターを被験者に投与し得る。

【０１５３】

更に別の実施態様では、限定するものではないが上に列記した疾患を含む、ＡＴＲＳ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ＡＴＲＳを有する組成物を好適な医薬用キャリアと共に被験者に投与し得る。

【０１５４】

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

【０１５５】
更なる実施態様では、ＡＴＲＳの発現または活性の増大に関連した疾患の治療または予防のために、被験者にＡＴＲＳのアンタゴニストを投与し得る。限定するものではないが、このような疾患の例には、上記した細胞増殖異常、自己免疫／炎症性の疾患が含まれる。一実施態様では、ＡＴＲＳ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ＡＴＲＳ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。

【０１５６】
別の実施態様では、限定するものではないが上に列記した疾患を含む、ＡＴＲＳ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、ＡＴＲＳをコードするポリヌクレオチドの相補配列を発現するベクターを被験者に投与し得る。

【０１５７】
別の実施態様では、本発明の任意のタンパク質、アンタゴニスト、抗体、アゴニスト、相補配列、またはベクターを、別の好適な治療薬と組み合わせ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。

【０１５８】
ＡＴＲＳ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し得る。 詳しくは、精製されたＡＴＲＳ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ＡＴＲＳと特異的に結合する薬の同定が可能である。ＡＴＲＳの抗体も、当分野で周知の方法を用いて産生し得る。このような抗体の例には、ポリクローナル抗体、モノクローナル抗体、キメラ抗体、一本鎖抗体、Ｆａｂ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（すなわち二量体の形成を阻害する抗体）は通常、治療用に好適である。

【０１５９】
抗体の産生のためには、ヤギ、ウサギ、ラット、マウス、ヒトその他など種々の宿主が、ＡＴＲＳまたは、免疫原性の特性を備えるその任意の断片またはオリゴペプチドの注入によって免疫化され得る。宿主の種に応じて、種々のアジュバントを用いて免疫応答を高めることもできる。限定するものではないがこのようなアジュバントには、フロイントアジュバントと、水酸化アルミニウムなどのミネラルゲルアジュバントと、リゾレシチン、プルロニックポリオル、ポリアニオン、ペプチド、油性乳剤、スカシガイのヘモシアニン（ＫＬＨ）、ジニトロフェノールなどの界面活性剤とがある。ヒトに用いられるアジュバントの中では、ＢＣＧ（カルメット‐ゲラン杆菌）およびコリネバクテリウム‐パルバム（Ｃｏｒｙｎｅｂａｃｔｅｒｉｕｍ ｐａｒｖｕｍ）が特に好ましい。

【０１６０】
ＡＴＲＳに対する抗体類を誘導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るアミノ酸配列を持つものが好ましく、一般的には約１０個以上のアミノ酸からなる。これらのオリゴペプチド、ペプチドまたは断片は、天然タンパク質のアミノ酸配列の一部と同一であることが望ましい。ＡＴＲＳ アミノ酸の短いストレッチは、ＫＬＨなど別のタンパク質の配列と融合されることができ、このキメラ分子に対する抗体群が産生され得る。

【０１６１】
ＡＴＲＳ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し得る。限定するものではないがこのような技術には、ハイブリドーマ技術、ヒトＢ細胞ハイブリドーマ技術、およびＥＢＶ－ハイブリドーマ技術がある（例えば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．Ａｃ

ａｄ．Ｓｃｉ．ＵＳＡ８０：２０２６－２０３０、Ｃｏｌｅ，Ｓ．Ｐ．他（１９８４）Ｍｏｌ．Ｃｅｌｌ Ｂｉｏｌ．６２：１０９－１２０を参照）。

【０１６２】

更に、「キメラ抗体」作製のために発達した、ヒト抗体遺伝子にマウス抗体遺伝子をスプライシングするなどの技術が、好適な抗原特異性およ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．ＵＳＡ８１：６８５１６８５５、Ｎｅｕｂｅｒｇｅｒ、Ｍ．Ｓ．他．（１９８４）Ｎａｔｕｒｅ ３１２：６０４－６０８、Ｔａｋｅｄａ，Ｓ．他（１９８５）Ｎａｔｕｒｅ ３１４：４５２－４５４を参照）。別法では、当分野で既知の方法を用いて、一本鎖抗体の産生のために記載された技術を適用して、ＡＴＲＳ特異的一本鎖抗体を生成し得る。関連した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類からチェーンシャッフリングによって産生することもできる（例えばＢｕｒｔｏｎ Ｄ．Ｒ．（１９９１）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ８８：１０１３４－１０１３７を参照）。

【０１６３】

抗体類の産生は、リンパ球集団における<u>ｉｎ ｖｉｖｏ</u>産生の誘導によって、あるいは免疫グロブリンライブラリ類のスクリーニングまたは文献に開示されているような、高度に特異的な結合試薬のパネル類のスクリーニングによっても行い得る（例えばＯｒｌａｎｄｉ，Ｒ．他（１９８９）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ８６：３８３３－３８３７、Ｗｉｎｔｅｒ，Ｇ．他（１９９１）Ｎａｔｕｒｅ ３４９：２９３－２９９を参照）。

【０１６４】

ＡＴＲＳに対する特異的な結合部位を持つ抗体断片をも産生し得る。例えば、限定するものではないが、このような断片には、抗体分子のペプシン消化によって産出される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産生されるＦａｂ断片とがある。あるいは、Ｆａｂ発現ライブラリを作製することによって、所望の特異性を持つモノクローナルＦａｂ断片を迅速且つ容易に同定することが可能となる（例えばＨｕｓｅ，Ｗ．Ｄ．他（１９８９）Ｓｃｉｅｎｃｅ ２４６：１２７５－１２８１を参照）。

【０１６５】

種々の免疫学的検定（イムノアッセイ）を用いてスクリーニングし、所望の特異性を有する抗体を同定することができる。確立された特異性を有するポリクローナル抗体またはモノクローナル抗体の何れかを用いる競合結合試験、または免疫放射定量測定法のための数々のプロトコルが、当分野では周知である。通常このようなイムノアッセイには、ＡＴＲＳとその特異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ＡＴＲＳエピトープに対して反応性を持つモノクローナル抗体群を用いる、２部位モノクローナルベースのイムノアッセイが一般に利用されるが、競合結合試験をも利用し得る（Ｐｏｕｎｄ、前出）。

【０１６６】

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、ＡＴＲＳに対する抗体の親和性を算定し得る。親和性を結合定数Ｋａで表すが、このＫａは、平衡状態下でＡＴＲＳ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。ポリクローナル抗体類は多数のＡＴＲＳエピトープに対して親和性が不均一であり、或るポリクローナル抗体試薬に関して判定したＫａは、ＡＴＲＳに対する抗体群の平均の親和性または結合活性を表す。或る特定ＡＴＲＳエピトープに単一特異的なモノクローナル抗体類の或る試薬について判定されたＫａは、親和性の真の測定値を表す。Ｋａ値が約10^9～10^{12}Ｌ／ｍｏｌの高親和性抗体試薬は、ＡＴＲＳ抗体複合体が激しい操作に耐えなければならないイムノアッセイに用いるのが好ましい。Ｋａ値が約10^6～10^7ｌｉｔｅｒ／ｍｏｌの低親和性抗体試薬は、ＡＴＲＳが抗体から最終的に（好ましくは活性化状態で）解離する必要がある免疫精製（ｉｍｍｕｎｏｐｕｒｉｆｉｃ

ａｔｉｏｎ）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）。

【０１６７】

ポリクローナル抗体試薬の抗体価およ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およ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の特異的な抗体を持つポリクローナル抗体試薬は一般に、ＡＴＲＳ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（例えば前出のＣａｔｔｙ、同Ｃｏｌｉｇａｎ 他を参照）。

【０１６８】

本発明の別の実施態様では、ＡＴＲＳをコードするポリヌクレオチド、またはその任意の断片や相補配列が、治療目的で使用され得る。ある実施態様では、ＡＴＲＳをコードする遺伝子のコーディング領域や調節領域に相補的な配列やアンチセンス分子（ＤＮＡ、ＲＮＡ、ＰＮＡ、または修飾されたオリゴヌクレオチド）を設計して遺伝子発現を変更できる。このような技術は当分野では周知であり、アンチセンスオリゴヌクレオチドまたはより大きな断片が、ＡＴＲＳをコードする配列の制御領域またはコード領域に沿ったさまざまな位置から設計され得る（例えばＡｇｒａｗａｌ，Ｓ．，編集（１９９６）Ａｎｔｉｓｅｎｓｅ Ｔｈｅｒａｐｅｕｔｉｃｓ，Ｈｕｍａｎａ Ｐｒｅｓｓ Ｉｎｃ．，Ｔｏｔａｗａ ＮＪを参照）。

【０１６９】

治療に用いる場合、アンチセンス配列を適切な標的細胞に導入するのに好適な、任意の遺伝子送達系を用いることができる。アンチセンス配列は、転写時に標的タンパク質をコードする細胞配列の少なくとも一部に相補的な配列を作製する発現プラスミドの形で、細胞内に送達し得る（例えばＳｌａｔｅｒ，Ｊ．Ｅ．他（１９９８）Ｊ．Ａｌｌｅｒｇｙ Ｃｌｉｎ．Ｉｍｍｕｎｏｌ．１０２（３）：４６９－４７５ およびＳｃａｎｌｏｎ，Ｋ．Ｊ．他（１９９５）９（１３）：１２８８－１２９６を参照）。アンチセンス配列はまた、例えばレトロウイルスやアデノ随伴ウイルスベクターなどのウイルスベクターを用いて細胞内に導入することもできる（例えばＭｉｌｌｅｒ，Ａ．Ｄ．（１９９０）Ｂｌｏｏｄ ７６：２７１、前出のＡｕｓｕｂｅｌ、Ｕｃｋｅｒｔ，Ｗ．およびＷ．Ｗａｌｔｈｅｒ（１９９４）Ｐｈａｒｍａｃｏｌ．Ｔｈｅｒ．６３（３）：３２３－３４７を参照）。その他の遺伝子送達機構には、リポソーム系、人工的なウイルスエンベロープおよび当分野で公知のその他のシステムが含まれる（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などを参照）。

【０１７０】

本発明の別の実施態様では、ＡＴＲＳ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得る。遺伝子治療により、（ｉ）遺伝子欠損症を治療し（例えばＸ染色体連鎖遺伝（Ｃａｖａｚｚａｎａ－Ｃａｌｖｏ，Ｍ．他（２０００）Ｓｃｉｅｎｃｅ ２８８：６６９－６７２）により特徴付けられる重症複合型免疫不全（ＳＣＩＤ）－Ｘ１病の場合）、先天性アデノシンデアミナーゼ（ＡＤＡ）欠損症に関連する重症複合型免疫不全症候群（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）Ｃｅｌ

ｌ７５：２０７－２１６：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や、第ＶＩＩＩ因子若しくは第ＩＸ因子欠損に起因する血友病（Ｃｒｙｓｔａｌ，Ｒ．Ｇ．（１９９５）Ｓｃｉｅｎｃｅ ２７０：４０４－４１０、Ｖｅｒｍａ，Ｉ．Ｍ．およびＮ．Ｓｏｍｉａ（１９９７）Ｎａｔｕｒｅ ３８９：２３９－２４２）、（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、Ｐｏｅｓｃｈｌａ，Ｅ．他（１９９６）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ．９３：１１３９５－１１３９９）などヒトレトロウイルス、Ｂ型若しくはＣ型肝炎ウイルス（ＨＢＶ、ＨＣＶ）、Ｃａｎｄｉｄａ ａｌｂｉｃａｎｓおよび Ｐａｒａｃｏｃｃｉｄｉｏｉｄｅｓ ｂｒａｓｉｌｉｅｎｓｉｓ等の寄生真菌、並びに熱帯熱マラリア原虫およびクルーズトリパノソーマ等の寄生原虫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。ＡＴＲＳの発現若しくは調節における遺伝子欠損が疾患を引き起こす場合、形質導入した細胞の好適な集団からＡＴＲＳを発現させて、その遺伝子欠損によって起こる症状の発現を緩和し得る。

【０１７１】

本発明の更なる実施態様では、ＡＴＲＳの欠損による疾患や障害は、ＡＴＲＳをコードする哺乳類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ＡＴＲＳ欠損細胞に導入することによって治療される。ｉｎ ｖｉｖｏあるいは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、および（ｖ）ＤＮＡトランスポゾンの使用がある（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）。

【０１７２】

ＡＴＲＳの発現に有効であり得る発現ベクターとしては、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。ＡＴＲＳを発現させるために、（ｉ）構成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－アクチンの遺伝子等）、（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．およびＨ．Ｂｕｊａｒｄ（１９９２）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．Ｕ．Ｓ．Ａ．８９：５５４７－５５５１；Ｇｏｓｓｅｎ，Ｍ．他（１９９５）Ｓｃｉｅｎｃｅ ２６８：１７６６－１７６９；Ｒｏｓｓｉ，Ｆ．Ｍ．Ｖ．およびＨ．Ｍ．Ｂｌａｕ（１９９８）Ｃｕｒｒ．Ｏｐｉｎ．Ｂｉｏｔｅｃｈｎｏｌ．９：４５１－４５６））、エクジソン誘導性プロモーター（市販されているプラスミドＰＶＧＲＸＲ及びＰＩＮＤに含まれている：Ｉｎｖｉｔｒｏｇｅｎ）、ＦＫ５０６／ラパマイシン誘導性プロモーター、またはＲＵ４８６／ミフェプリストーン誘導性プロモーター（Ｒｏｓｓｉ，Ｆ．Ｍ．Ｖ．およびＨ．Ｍ．Ｂｌａｕ，前出）、または（ｉｉｉ）正常な個体に由来するＡＴＲＳをコードする内因性遺伝子の天然プロモーター若しくは組織特異的プロモーターを用い得る。

【０１７３】

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、当業者は実験の各パラメータを最適化する努力をさほど要さず、ポリヌクレオチド群を、培養中の標的細胞群に送達し得る。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、標準化された哺乳類の形質移入プロトコルの修正が必要である。

【０１７４】

本発明の別の実施態様では、ＡＴＲＳの発現に関連する遺伝子欠損によって起こる疾患や障害が、（ｉ）レトロウイルス長末端反復配列（ＬＴＲ）プロモーター若しくは独立プロモーターのコントロール下でＡＴＲＳ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配列と効率的なベクター増殖に必要なコーディング配列とを伴う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される。レトロウイルスベクター（例えばＰＦＢおよび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．ＵＳＡ ９２：６７３３－６７３７）に基づく。上記データを引用することを以て本明細書の一部とする。このベクターは、好適なベクター産生細胞株（ＶＰＣＬ）において増殖される。ＶＰＣＬは、各標的細胞上の受容体への親和性を持つエンベロープ遺伝子を、またはＶＳＶｇなど汎親和性エンベロープタンパク質を発現する（Ａｒｍｅｎｔａｎｏ，Ｄ．他（１９８７）Ｊ．Ｖｉｒｏｌ．６１：１６４７－１６５０、Ｂｅｎｄｅｒ，Ｍ．Ａ．他（１９８７）Ｊ．Ｖｉｒｏｌ．６１：１６３９－１６４６、Ａｄａｍ，Ｍ．Ａ．およびＡ．Ｄ．Ｍｉｌｌｅｒ（１９８８）Ｊ．Ｖｉｒｏｌ．６２：３８０２－３８０６、Ｄｕｌｌ，Ｔ．他（１９９８）Ｊ．Ｖｉｒｏｌ．７２：８４６３－８４７１、Ｚｕｆｆｅｒｅｙ，Ｒ．他（１９９８）Ｊ．Ｖｉｒｏｌ．７２：９８７３－９８８０）。Ｒｉｇｇに付与された米国特許第５，９１０，４３４号（「Ｍｅｔｈｏｄ ｆｏｒ ｏｂｔａｉｎｉｎｇ ｒｅｔｒｏｖｉｒｕｓ ｐａｃｋａｇｉｎｇ ｃｅｌｌ ｌｉｎｅｓ ｐｒｏｄｕｃｉｎｇ ｈｉｇｈ ｔｒａｎｓｄｕｃｉｎｇ ｅｆｆｉｃｉｅｎｃｙ ｒｅｔｒｏｖｉｒａｌ ｓｕｐｅｒｎａｔａｎｔ」）において、レトロウイルスパッケージング細胞株群を得る方法が開示されており、引用を以て本明細書の一部とする。レトロウイルスベクター類の増殖や、細胞集団（例えばＣＤ４$^{+}$Ｔ細胞群）の形質導入、および形質導入した細胞群の患者への戻しは、遺伝子治療の分野では当業者に周知の手法であり、多数の文献に記載がある（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）。

【０１７５】

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、ＡＴＲＳの発現に関連する１つ以上の遺伝子異常を有する細胞に、ＡＴＲＳをコードするポリヌクレオチドを送達する。アデノウイルス系ベクター類の作製およびパッケージングについては、当業者に周知である。複製欠損型アデノウイルスベクター類は、種々の免疫調節タンパク質をコードする遺伝子群を、無損傷の膵島内にインポートする目的で多様に利用し得ることが証明された（Ｃｓｅｔｅ，Ｍ．Ｅ．他（１９９５）Ｔｒａｎｓｐｌａｎｔａｔｉｏｎ ２７：２６３－２６８）。潜在的に有用なアデノウイルスベクター類は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、引用することを以て本明細書の一部とする。アデノウイルスベクター類については、Ａｎｔｉｎｏｚｚｉ，Ｐ．Ａ．他（

１９９９）Ａｎｎｕ．Ｒｅｖ．Ｎｕｔｒ．１９：５１１－５４４並びに、Ｖｅｒｍａ，Ｉ．Ｍ．およびＮ．Ｓｏｍｉａ（１９９７）Ｎａｔｕｒｅ１８：３８９：２３９－２４２も参照されたい。両文献は、引用を以て本明細書の一部とする。

【０１７６】

別法では、ヘルペス系遺伝子治療の送達系を用いて、ＡＴＲＳの発現に関連する１つ以上の遺伝子異常を有する標的細胞に、ＡＴＲＳ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）系ベクターの使用は、ＨＳＶが向性を持つ中枢神経細胞にＡＴＲＳを導入する際に特に有用であり得る。ヘルペス系ベクター類の作製およびパッケージングは、当業者に公知である。或る複製適格性単純ヘルペスウイルス（ＨＳＶ）Ｉ型系のベクターが、或るレポーター遺伝子の、霊長類の眼への送達に用いられている（Ｌｉｕ，Ｘ．他（１９９９）Ｅｘｐ．Ｅｙｅ Ｒｅｓ．１６９：３８５－３９５）。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 ｓｔｒａｉｎｓ ｆｏｒ ｇｅｎｅ ｔｒａｎｓｆｅｒ」）に詳細に開示されており、該特許の引用を以て本明細書の一部とする。米国特許第５，８０４，４１３号は、ヒト遺伝子治療などの目的で、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つの外来遺伝子を有するゲノムを持つ、組換えＨＳＶ ｄ９２の利用について記載する。上記特許はまた、ＩＣＰ４、ＩＣＰ２７、およびＩＣＰ２２を欠失した組換えＨＳＶ系統の、作製および使用について開示している。ＨＳＶベクター類については、Ｇｏｉｎｓ，Ｗ．Ｆ．他（１９９９）Ｊ．Ｖｉｒｏｌ．７３：５１９－５３２およびＸｕ，Ｈ．他（１９９４）Ｄｅｖ．Ｂｉｏｌ．１６３：１５２－１６１も参照されたい。両文献は、引用を以て本明細書の一部とする。クローン化したヘルペスウイルス配列群の操作、大ヘルペスウイルスゲノム（ｌａｒｇｅ ｈｅｒｐｅｓｖｉｒｕｓ ｇｅｎｏｍｅｓ）の様々なセグメントを持つ複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および増殖、並びに細胞群へのヘルペスウイルスの感染は、当業者に公知の技術である。

【０１７７】

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ＡＴＲＳ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（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類は、ＳＦＶゲノムに基づく（Ｇａｒｏｆｆ，Ｈ．およびＫ．－Ｊ．Ｌｉ（１９９８）Ｃｕｒｒ．Ｏｐｉｎ．Ｂｉｏｔｅｃｈｎｏｌ．９：４６４－４６９）。αウイルスＲＮＡの複製中に、ウイルスのカプシドタンパク質群を通常はコードする、サブゲノムＲＮＡが産生される。このサブゲノムＲＮＡは、完全長ゲノムＲＮＡよりも高レベルに複製するので、酵素活性（例えばプロテアーゼおよびポリメラーゼ）を持つウイルスタンパク質に比べてカプシドタンパク質群が過剰産生される。同様に、ＡＴＲＳをコードする配列をαウイルスゲノムのカプシドをコードする領域に挿入することによって、ベクター形質導入細胞においてＡＴＲＳをコードする多数のＲＮＡが産生され、高いレベルでＡＴＲＳが合成される。通常、αウイルスの感染は、数日以内の細胞溶解を伴う。一方、シンドビスウイルス（ＳＩＮ）の或る変異体を有するハムスター正常腎臓細胞（ＢＨＫ－２１）群が持続的な感染を確立する能力は、αウイルス類の溶解複製を、遺伝子治療に応用し得るように好適に改変可能であることを示唆する（Ｄｒｙｇａ，Ｓ．Ａ．他（１９９７）Ｖｉｒｏｌｏｇｙ ２２８：７４－８３）。広範な宿主にαウイルスを導入できるので、様々なタイプの細胞にＡＴＲＳを導入できることとなる。或る集団における或るサブセットの細胞群の特異的形質導入は、形質導入前に細胞の選別を必要とし得る。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およびＲＮＡの形質移入方法およびαウイルスの感染方法は、当業者に公知である。

【０１７８】

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。この位置は、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、二重らせんの能力を阻害するので有用である。三重らせんＤＮＡを用いる最近の治療の進歩については文献に記載がある（例えばＧｅｅ，Ｊ．Ｅ．他（１９９４）ｉｎ Ｈｕｂｅｒ，Ｂ．Ｅ．およびＢ．Ｉ．Ｃａｒｒ，Ｍｏｌｅｃｕｌａｒ ａｎｄ Ｉｍｍｕｎｏｌｏｇｉｃ Ａｐｐｒｏａｃｈｅｓ，Ｆｕｔｕｒａ Ｐｕｂｌｉｓｈｉｎｇ，Ｍｔ．Ｋｉｓｃｏ ＮＹ，１６３－１７７ページを参照）。また、相補配列またはアンチセンス分子を設計し、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の翻訳をブロックし得る。

【０１７９】

リボザイム類は、酵素性ＲＮＡ分子である。ＲＮＡの特異的切断を触媒するためにリボザイムを用い得る。リボザイム作用のメカニズムは、相補的標的ＲＮＡへのリボザイム分子の配列特異的ハイブリダイゼーションと、その後に起こる内ヌクレオチド鎖切断に関与している。例えば、人為操作されたハンマーヘッド型リボザイム分子は、ＡＴＲＳをコードする配列の内ヌクレオチド鎖切断を、特異的且つ効果的に触媒する可能性がある。

【０１８０】

任意の潜在的ＲＮＡ標的内の特異的リボザイム切断部位を先ず同定するため、ＧＵＡ、ＧＵＵ、ＧＵＣ配列などリボザイム切断部位に対して標的分子をスキャンする。一度同定されると、切断部位を持つ標的遺伝子の領域に対応する１５～２０リボヌクレオチドの短いＲＮＡ配列が、そのオリゴヌクレオチドを機能不全にするような２次構造の特徴をもっていないかを評価することが可能になる。候補標的の適合性の評価も、リボヌクレアーゼ保護アッセイを用いて相補的オリゴヌクレオチド群とのハイブリダイゼーションへのアクセス可能性をテストすることによって行い得る。

【０１８１】

本発明の相補リボ核酸分子およびリボザイムは、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。作製方法には、固相ホスホラミダイト化学合成など、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、ＡＴＲＳをコードするＤＮＡ配列のｉｎ ｖｉｔｒｏ及びｉｎ ｖｉｖｏ転写によってＲＮＡ分子を産生し得る。このようなＤＮＡ配列は、Ｔ７やＳＰ６などの好適なＲＮＡポリメラーゼプロモーターを用いて多様なベクター内に取り込むことが可能である。あるいは、相補的ＲＮＡを構成的あるいは誘導的に合成するこれらのｃＤＮＡ構成物を、細胞株、細胞または組織内に導入することができる。

【０１８２】

細胞内の安定性を高め、半減期を長くするためにＲＮＡ分子を修飾し得る。限定するものではないが可能な修飾としては、分子の５’末端、３’末端、あるいはその両方において隣接配列群を追加することや、分子の主鎖内においてホスホジエステラーゼ結合ではなくホスホロチオエートまたは２’Ｏ－メチルを使用することが含まれる。この概念は、本来はＰＮＡ群の産出におけるものであるが、これら全ての分子に拡大することができる。そのためには、内因性エンドヌクレアーゼによって容易には認識されない、アデニン、シチジン、グアニン、チミン、およびウリジンにアセチル－、メチル－、チオ－および同様の修飾をしたものや、非従来型塩基、例えばイノシン、クエオシン（ｑｕｅｏｓｉｎｅ）、ワイブトシン（ｗｙｂｕｔｏｓｉｎｅ）を加える。

【０１８３】

本発明の更なる実施態様は、ＡＴＲＳをコードするポリヌクレオチドの発現の改変に有効な化合物をスクリーニングする方法を含む。限定するものではないが特異ポリヌクレオチドの発現改変を起こすのに有効であり得る化合物には、オリゴヌクレオチド、アンチセンスオリゴヌクレオチド、三重らせん形成オリゴヌクレオチドや、転写因子などのポリペプチド転写制御因子、および、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改変し得る。従って、ＡＴＲＳ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、ＡＴＲＳ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得る。また、ＡＴＲ

Ｓ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、ＡＴＲＳ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。

【０１８４】

或る特定ポリヌクレオチドの発現を改変する際の有効性に対して、少なくとも１個または複数個の試験化合物をスクリーニングし得る。試験化合物を得るには、当分野で公知の任意の方法を用い得る。取得方法としては、以下の場合に有効な既知化合物の化学修飾がある。ポリヌクレオチドの発現を改変する場合と、既存の、商用または専用の、天然または非天然の化合物の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および／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とである。ＡＴＲＳをコードするポリヌクレオチドを有するサンプルは、このようにして得た少なくとも１つの試験化合物に曝露される。サンプルは例えば、無傷細胞、透過化処理した細胞、ｉｎ ｖｉｔｒｏ無細胞系すなわち再構成生化学系があり得る。ＡＴＲＳをコードするポリヌクレオチドの発現における改変は、当分野で公知の任意の方法でアッセイする。通常、ＡＴＲＳをコードするポリヌクレオチドの配列に相補的なヌクレオチド配列を有するプローブを用いたハイブリダイゼーションにより、特定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することにより、１つ以上の試験化合物に曝露される、または曝露されないポリヌクレオチドの発現の比較のための基礎を形成し得る。試験化合物に曝露されるポリヌクレオチドの発現における変化の検出は、ポリヌクレオチドの発現を改変する際に試験化合物が有効であることを示す。或る特定ポリヌクレオチドの改変発現に有効な化合物のためのスクリーニングを実行でき、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細胞等のヒト細胞株（Ｃｌａｒｋｅ，Ｍ．Ｌ．他（２０００）Ｂｉｏｃｈｅｍ．Ｂｉｏｐｈｙｓ．Ｒｅｓ．Ｃｏｍｍｕｎ．２６８：８－１３）を用いる。本発明の或る特定の実施例は、或る特定ポリヌクレオチド配列に対するアンチセンス活性について、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、修飾した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する（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）。

【０１８５】

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および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患者に自家移植で戻すことができる。トランスフェクションによる、またはリボソーム注入やポリカチオンアミノポリマーによる送達は、当分野でよく知られている方法を用いて実行することができる（Ｇｏｌｄｍａｎ，Ｃ．Ｋ．他（１９９７）Ｎａｔ．Ｂｉｏｔｅｃｈｎｏｌ．１５：４６２－４６６などを参照）。

【０１８６】

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、サルなどの哺乳類を含めて治療が必要な全ての被験体に適用できる。

【０１８７】

本発明の或る更なる実施例は、薬剤として許容できる賦形剤と共に処方される活性成分を一般に持つ、或る組成物の投与に関する。賦形剤には例えば、糖、でんぷん、セルロース、ゴムおよびタンパク質がある。様々な処方が広く知られており、詳細はＲｅｍｉｎｇｔｏｎ’ｓ Ｐｈａｒｍａｃｅｕｔｉｃａｌ Ｓｃｉｅｎｃｅｓ（Ｍａａｃｋ Ｐｕｂｌｉｓｈｉｎｇ，Ｅａｓｔｏｎ ＰＡ）の最新版に記載されている。このような組成物は、ＡＴＲＳ、ＡＴＲＳの抗体、擬態物質、アゴニスト、アンタゴニスト、またはインヒビターなどを有し得る。

【０１８８】
本発明に用いる組成物は、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。
【０１８９】
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およびタンパク質）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、インスリンなどの薬剤を実質的に血液循環へ輸送することを可能にした（Ｐａｔｔｏｎ，　Ｊ．Ｓ．　他，　米国特許第５，９９７，８４８号などを参照）。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。
【０１９０】
本発明での使用に適した組成物には、所定の目的を達成するために必要なだけの量の活性成分を含有する組成物が含まれる。有効投与量の決定は、当業者の能力の範囲内で行う。
【０１９１】
ＡＴＲＳまたはその断片を有する高分子群を直接に細胞内へ送達するべく、特殊な形状の組成物が調製され得る。例えば、細胞不透過性高分子を含むリポソーム製剤は、その高分子の細胞融合と細胞内送達とを促進し得る。別法では、ＡＴＲＳまたはその断片を、ＨＩＶ　Ｔａｔ－１タンパク質の短いカチオンＮ末端部に結合することもできる。このようにして産生された融合タンパク質類は、或るマウスモデル系の、脳を含む全ての組織の細胞群に形質導入することがわかっている（Ｓｃｈｗａｒｚｅ，　Ｓ．Ｒ．　他　（１９９９）　Ｓｃｉｅｎｃｅ　２８５：１５６９－１５７２）。
【０１９２】
任意の化合物に対して、先ず細胞培養アッセイ、例えば新生物性細胞の細胞培養アッセイにおいて、あるいは、動物モデル、例えばマウス、ラット、ウサギ、イヌ、サルまたはブタなどにおいて、治療有効量を推定できる。動物モデルはまた、好適な濃度範囲および投与経路を決定するためにも用い得る。このような情報を用いて、次にヒトに対する有益な投与量および投与経路を決定することができる。
【０１９３】
治療有効量とは、症状や容態を回復させる、例えばＡＴＲＳまたはその断片、ＡＴＲＳの抗体、アゴニストまたはアンタゴニスト、インヒビターなど、活性処方成分の量を指す。治療有効度および毒性は、細胞培養または動物実験における標準的な薬学手法によって、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）統計を計算するなどして判定できる。毒性効果の、治療効果に対する投与量比が治療指数であり、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。細胞培養アッセイと動物実験とから得られたデータは、ヒトに用いる投与量の範囲での調剤に用いられる。このような組成物に含まれる投与量は、毒性を殆どあるいは全く持たず、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、患者の感受性および投与の経路によって、投与量はこの範囲内で様々に変わる。
【０１９４】
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。充分なレベルの活性成分を与え、あるいは所望の効果を維持すべく、用法および用量を調整する。被験者に関する要素としては、疾患の重症度、患者の一般的な健康状態、患者の年齢、体重および性別、投与の時間および頻度、薬剤の配合、反応感受性および治療に対する応答などを考慮し得る。作用期間が長い組成物は、特定の製剤の半減期およ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、あるいは２週間に１度の間隔で投与し得る。

【０１９５】
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。特定の投与量およ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、ヌクレオチドの処方では、タンパク質またはそれらのインヒビター類とは異なる処方を利用することになる。同様に、ポリヌクレオチドまたはポリペプチドの送達は、特定の細胞、症状、部位などに特異的なものとなる。

【０１９６】
（診断）
別の実施態様では、ＡＴＲＳに特異的に結合する抗体が、ＡＴＲＳの発現によって特徴付けられる障害の診断に、またはＡＴＲＳやそ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中の患者をモニターするためのアッセイに用いられ得る。診断目的に有用な抗体は、上記の治療の箇所で記載した方法と同じ方法で調合される。ＡＴＲＳの診断アッセイには、抗体及び標識を用いて、ヒトの体液中、或いは細胞や組織の抽出物中のＡＴＲＳ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。

【０１９７】
ＡＴＲＳを測定するための、ＥＬＩＳＡ，ＲＩＡ，及びＦＡＣＳなど種々のプロトコルは、当分野では既知であり、変容した或いは異常なレベルのＡＴＲＳ発現を診断するための基盤を提供する。正常或いは標準的なＡＴＲＳ発現の値は、複合体の形成に適した条件下で、正常な哺乳動物、例えばヒトなどの被験体から採取した体液または細胞抽出物と、ＡＴＲＳに対する抗体とを混合することによって確定する。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。被験体、対照及び、生検組織からの疾患サンプルでのＡＴＲＳの発現の量が、この標準値と比較される。標準値と被験体との偏差が、疾患を診断するパラメータを確定する。

【０１９８】
本発明の別の実施態様では、ＡＴＲＳをコードするポリヌクレオチドを診断のために用い得る。用い得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および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、ＡＴＲＳの発現が疾患と相関し得る生検組織における遺伝子発現を検出し定量し得る。この診断アッセイを用いて、ＡＴＲＳの有無や、過剰な発現を調べ得る。また、治療中のＡＴＲＳレベルの調節を監視し得る。

【０１９９】
一実施形態では、ＡＴＲＳまたは近縁の分子をコードする、ゲノム配列など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、ＡＴＲＳをコードする核酸配列を同定し得る。プローブが高度に特異的な領域（例えば５’調節領域）から作られている、或いはやや特異性の低い領域（例えば保存されたモチーフ）から作られているかにかかわらず、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シーとによって、そのプローブがＡＴＲＳをコードする天然配列のみを同定するかどうか、或いは対立遺伝子変異配列や関連配列を同定するかどうかが決まることとなる。

【０２００】
プローブはまた、関連する配列の検出に利用され得る他、ＡＴＲＳをコードする任意の配列との少なくとも５０％の配列同一性を有し得る。当該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：２の配列に、或いはＡＴＲＳ遺伝子の、プロモーター、エンハンサー、イントロンなど、ゲノム配列に由来し得る。

【０２０１】

ＡＴＲＳ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、ＡＴＲＳまたはＡＴＲＳ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プローブ作製用ベクター類は、当業者に知られており、市販されており、好適なＲＮＡポリメラーゼと好適な標識されたヌクレオチドと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得る。ハイブリダイゼーションプローブ類は、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などの放射性核種が、あるいはアビジン／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など酵素標識が挙げられる。

【０２０２】

ＡＴＲＳ をコードするポリヌクレオチド配列を用いて、ＡＴＲＳの発現に関連する疾患を診断し得る。限定するものではないが、このような疾患には、細胞増殖異常として、日光角化症、動脈硬化、アテローム性硬化、滑液包炎、硬変、肝炎、混合性結合組織病（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、また腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫及び奇形癌など癌、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、子宮の癌があり、自己免疫／炎症性疾患には、後天性免疫不全症候群（ＡＩＤＳ）、アジソン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血症、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫性多腺性内分泌カンジダ症外胚葉ジストロフィ（ＡＰＥＣＥＤ）、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー皮膚炎、皮膚筋炎、真性糖尿病、肺気腫、リンパ球傷害因子性偶発性リンパ球減少症、新生児溶血性疾患（胎児赤芽球症）、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性腸症候群、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、慢性関節リウマチ、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、ぶどう膜炎、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、外傷が含まれる。ＡＴＲＳをコードするポリヌクレオチド配列は、変容したＡＴＲＳ発現を検出するために患者から採取した体液或いは組織を利用する、サザ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や、ＰＣＲ法や、ディップスティック（ｄｉｐｓｔｉｃｋ）、ピン（ｐｉｎ）およびマルチフォーマットのＥＬＩＳＡ式アッセイ、及びマイクロアレイに使用し得る。このような定性方法または定量方法は、当分野で公知である。

【０２０３】

或る実施態様では、ＡＴＲＳ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。ＡＴＲＳ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。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化した場合は、サンプル内の、ＡＴＲＳをコードするヌクレオチド配列のレベル変化の存在が、関連する疾患の存在を標示する。このようなアッセイは、動物実験、臨床試験における特定の治療効果を推定するため、あるいは個々の患者の治療をモニターするために用いることもできる。

【０２０４】

ＡＴＲＳの発現に関連する疾患の診断の基盤を提供するために、発現の正常すなわち標準的なプロフィールが確立される。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体から採取された体液或いは細胞抽出物と、ＡＴＲＳ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で用いて行った実験から得た値を正常な被験体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。

【０２０５】

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

【０２０６】

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

【０２０７】

ＡＴＲＳ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、更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、酵素により産生するか、あるいはｉｎ ｖｉｔｒｏで産出し得る。オリゴマーは、好ましくは、ＡＴＲＳをコードするポリヌクレオチドの断片、或いはＡＴＲＳをコードするポリヌクレオチドに相補的なポリヌクレオチドの断片を有し、最適化した条件下で、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シー条件下で、近縁のＤＮＡあるいはＲＮＡ配列の検出、定量、あるいはその両方のため用いることが可能である。

【０２０８】

或る実施態様では、ＡＴＲＳ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、挿入および欠失である。限定するものではないがＳＮＰの検出方法には、ＳＳＣＰ（ｓｉｎｇｌｅ－ｓｔｒａｎｄｅｄ ｃｏｎｆｏｒｍａｔｉｏｎ ｐｏｌｙｍｏｒｐｈｉｓｍ）および蛍光ＳＳＣＰ（ｆＳＳＣＰ）がある。ＳＳＣＰでは、ＡＴＲＳ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、体液などに由来し得る。ＤＮＡ内のＳＮＰは、一本鎖形のＰＣＲ生成物の２次および３次構造に差異を生じさせる。差異は非変性ゲル中でのゲル電気泳動法を用いて検出可能である。ｆＳＳＣＰでは、オリゴヌクレオチドプライマーを蛍光的に標識する。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断片群の配列を比較することにより、多型性を同定し得る。これらのコンピュータベースの方法は、ＤＮＡの実験室での調製に、または統計モデルとＤＮＡ配列クロマトグラムの自動分析とを用いたシークエンシングのエラーに起因する配列変異を、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。

【０２０９】

ＡＴＲＳ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、対照核酸の共増幅（ｃｏａｍｐｌｉｆｉｃａｔｉｏｎ）及び、標準曲線から得た結果の補間もある（例えば、Ｍｅｌｂｙ，Ｐ．Ｃ．他（１９９３）Ｊ．Ｉｍｍｕｎｏｌ．Ｍｅｔｈｏｄｓ，１５９：２３５－２４４；Ｄｕｐｌａａ，Ｃ．他（１９９３）

Ａｎａｌ．　Ｂｉｏｃｈｅｍ．　２１２：２２９－２３６を参照）。複数サンプルの定量速度を加速するには、目的のオリゴマーまたはポリヌクレオチドを種々の希釈液中に置き、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でのアッセイを行い得る。

【０２１０】

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に由来するオリゴヌクレオチドまたはより長い断片を、或るマイクロアレイにおけるエレメント群として用いることができる。多数の遺伝子の相対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、遺伝子の変異体、突然変異および多型性の同定に用いることができる。この情報を用いることで、遺伝子機能を決定し、疾患の遺伝的根拠を理解し、疾患を診断し、遺伝子発現の機能としての疾病の進行／後退をモニターし、疾病治療における薬剤の活性を開発およびモニターすることができる。特に、患者にとって最もふさわしく、有効な治療法を選択するために、この情報を用いて患者の薬理ゲノムプロフ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効果的で副作用を殆ど示さない治療薬を選択することができる。

【０２１１】

別の実施態様では、ＡＴＲＳ、その断片または、ＡＴＲＳに特異的な抗体を、１マイクロアレイ上のエレメント群として用い得る。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－標的相互作用およ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。

【０２１２】

或る実施態様は、或る組織または細胞タイプの転写イメージを作製する、本発明のポリヌクレオチドの使用に関連する。転写イメージは、特定の組織または細胞タイプによる、遺伝子発現の包括的パターンを表す。包括的遺伝子発現パターンは、所与の条件下で所与の時間に発現した遺伝子の数および相対存在量を定量することにより分析し得る（Ｓｅｉｌｈａｍｅｒ 他の米国特許第５，８４０，４８４号「Ｃｏｍｐａｒａｔｉｖｅ Ｇｅｎｅ Ｔｒａｎｓｃｒｉｐｔ Ａｎａｌｙｓｉｓ」を参照。該特許は特に引用することを以って本明細書の一部となす）。したがって、特定の組織または細胞タイプの転写物または逆転写物全体に、本発明のポリヌクレオチドまたはそれらの相補体をハイブリダイズすることにより、転写イメージを作製し得る。或る実施態様では、本発明のポリヌクレオチドまたはそれらの相補体が１マイクロアレイ上に複数エレメントの１サブセットを持つ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。

【０２１３】

転写イメージは、組織、細胞株、生検またはその生体サンプルから単離した転写物を用いて作製し得る。転写イメージはしたがって、組織または生検サンプルの場合にはｉｎ ｖｉｖｏ、細胞株の場合にはｉｎ ｖｉｔｒｏでの遺伝子発現を反映する。

【０２１４】

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、ｉｎ ｖｉｔｒｏモデル系および薬剤の前臨床評価に、あるいは工業的または天然の環境化合物の毒性試験に関連して使用し得る。全ての化合物は、作用および毒性のメカニズムを標示し、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、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、該文献は特に引用を以って本明細書の一部となす）。試験化合物が、既知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、多数の遺伝子および遺伝子ファミリーからの発現情報を含んでいる場合に、最も有用且つ正確である。理想的には、ゲノム

全域にわたる発現の測定が、最高品質のシグネチャを提供する。たとえ、発現が任意の試験された化合物によって変容しない遺伝子があったとしても、それらの発現レベルを残りの発現データを正規化することに使用できるため、それらの遺伝子は重要である。正規化手順は、種々の化合物で処理した後の発現データの比較に有用である。或る毒性シグネチャのエレメント群に遺伝子機能を割り当てることは毒性機序の解釈に役立つが、毒性の予測につながる、シグネチャ群の統計的マッチングには、遺伝子機能の知識は必要でない（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）。したがって、毒性シグネチャを用いる中毒学的スクリーニングの際に、全ての発現した遺伝子配列を含めることは、重要且つ望ましい。

【０２１５】

或る実施例では、核酸を有する生体サンプルを試験化合物で処理することにより、この試験化合物の毒性を算定する。処理した生体サンプル中で発現した核酸は、本発明のポリヌクレオチドに特異的な１つ以上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。処理した生体サンプル中の転写レベルを、未処理生体サンプル中のレベルと比較する。両サンプルの転写レベルの差は、処理済サンプル中で試験化合物が引き起こす毒性反応を標示する。

【０２１６】

別の実施態様は、本発明のポリペプチド配列群を用いて或る組織または細胞タイプのプロテオームを分析することに関する。プロテオームの語は、特定の組織または細胞タイプでのタンパク質発現の包括的パターンを指す。プロテオームの各タンパク質成分は、個々に更なる分析の対象とすることができる。プロテオーム発現パターンすなわちプロフィールは、所与の条件下で所与の時間に発現したタンパク質の数および相対存在量を定量することにより分析し得る。したがって、或る細胞のプロテオームのプロフィールは、特定の組織または細胞タイプのポリペプチドを分離およ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、通常はクーマシーブルー、あるいは銀染色液または蛍光染色液などの物質を用いてゲルを染色することにより、分散した、独自の位置にある点としてゲル中で可視化される。各タンパク質スポットの光学密度は、通常、サンプル中のタンパク質レベルに比例する。異なるサンプル、例えば試験化合物または治療薬で処理または未処理のいずれかの生体サンプルからの、同等に位置した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。或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データが得られる。

【０２１７】

プロテオームのプロフィールは、ＡＴＲＳに特異的な抗体を用いてＡＴＲＳ発現レベルを定量することによっても作成し得る。或る実施態様では、マイクロアレイ上のエレメントとしてこれら抗体を用い、マイクロアレイをサンプルに曝して各アレイエレメントへのタンパク質結合レベルを検出することにより、タンパク質発現レベルを定量する（Ｌｕｅｋｉｎｇ，Ａ．他（１９９９）Ａｎａｌ．Ｂｉｏｃｈｅｍ．２７０：１０３－１１１、Ｍｅｎｄｏｚｅ，Ｌ．Ｇ．他（１９９９）Ｂｉｏｔｅｃｈｎｉｑｕｅｓ ２７：７７８－７８８）。検出は当分野で既知の様々な方法で行うことができ、例えば、チオール反応性またはアミノ反応性蛍光化合物とサンプル中のタンパク質を反応させ、各アレイエ

レメントにおける蛍光結合の量を検出し得る。

【０２１８】

プロテオームレベルでの毒性シグネチャも中毒学的スクリーニングに有用であり、転写レベルでの毒性シグネチャと並行に分析するべきである。いくつかの組織のいくつかのタンパク質については、転写物の存在量とタンパク質の存在量との相関が乏しいので（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改変するような化合物の分析において、プロテオーム毒性シグネチャは有用たり得る。更に、体液中の転写物の分析はｍＲＮＡの急速な分解のために困難なので、プロテオームのプロフィール作成はこのような場合により信頼し得、情報価値があり得る。

【０２１９】

別の実施例では、タンパク質を含有する生体サンプルを試験化合物で処理することにより試験化合物の毒性を算定する。処理された生体サンプル中で発現したタンパク質は、各タンパク質の量を定量し得るように分離する。各タンパク質の量を、未処理生体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の量の差は、処理サンプル中の試験化合物に対する毒性反応を標示する。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。

【０２２０】

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

【０２２１】

マイクロアレイは、本技術分野で既知の方法で調製し、使用し、分析する（例えばＢｒｅｎｎａｎ，Ｔ．Ｍ．他（１９９５）米国特許第５，４７４，７９６号、Ｓｃｈｅｎａ，Ｍ．他（１９９６）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ ９３：１０６１４－１０６１９、Ｂａｌｄｅｓｃｈｗｅｉｌｅｒ他（１９９５）ＰＣＴ出願第ＷＯ９５／２５１１１６号、Ｓｈａｌｏｎ，Ｄ．他（１９９５）ＰＣＴ出願第ＷＯ９５／３５５０５号、Ｈｅｌｌｅｒ，Ｒ．Ａ．他（１９９７）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ ９４：２１５０－２１５５、Ｈｅｌｌｅｒ，Ｍ．Ｊ．他（１９９７）米国特許第５，６０５，６６２号を参照）。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に記載がある。該文献は、特に引用を以って本明細書の一部となす。

【０２２２】

本発明の別の実施態様ではまた、ＡＴＲＳをコードする核酸配列を用いて、天然ゲノム配列をマッピングするのに有用なハイブリダイゼーションプローブを産生し得る。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、いくつかの例では、コード配列より非コード配列が好ましい。例えば、多重遺伝子族のメンバー内でのコード配列の保存により、染色体マッピング中に望ましくないクロスハイブリダイゼーションが生じる可能性がある。配列は、或る特定の染色体に、または或る染色体の或る特定領域に、または人為形成の染色体、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、あるいは単一染色体ｃＤＮＡライブラリ群に対してマッピングされる（例えばＨａｒｒｉｎｇｔｏｎ，Ｊ．Ｊ．他（１９９７）Ｎａｔ．Ｇｅｎｅｔ．１５：３４５－３５５；Ｐｒｉｃｅ，Ｃ．Ｍ．（１９９３）Ｂｌｏｏｄ

Ｒｅｖ．７：１２７－１３４；Ｔｒａｓｋ，Ｂ．Ｊ．（１９９１）Ｔｒｅｎｄｓ Ｇｅｎｅｔ．７：１４９－１５４を参照）。一度マッピングすると、本発明の核酸配列群を用いて、例えば或る病状の遺伝を特定染色体領域の遺伝と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開発し得る（例えば、Ｌａｎｄｅｒ，Ｅ．Ｓ．およびＤ．Ｂｏｔｓｔｅｉｎ（１９８６）Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ ８３：７３５３－７３５７を参照）。

【０２２３】

蛍光原位置ハイブリッド形成法（ＦＩＳＨ）は、他の物理的および遺伝的地図データと相関し得る（例えば前出のＨｅｉｎｚ－Ｕｌｒｉｃｈ，他（１９９５）ｉｎ Ｍｅｙｅｒｓ，９６５－９６８ページを参照）。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られる。物理的染色体地図上の、ＡＴＲＳ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。

【０２２４】

確定した染色体マーカー類を用いた連鎖分析などの物理的マッピング技術、および染色体標本の原位置ハイブリッド形成法を用いて、遺伝地図を拡張することができる。例えばマウスなど別の哺乳類の染色体上に遺伝子を配置することにより、正確な染色体上の遺伝子座が未知でも、関連するマーカー類をしばしば明らかにし得る。この情報は、位置クローニングなどの遺伝子発見技術を用いて疾患遺伝子を探る研究者にとって価値がある。いったん疾患または症候群に関与する遺伝子（群）が、血管拡張性失調症の１１ｑ２２－２３領域など、特定のゲノム領域への遺伝的連鎖によって、大まかに位置決めがなされると、該領域にマップされる任意の配列は、更なる調査のための関連遺伝子あるいは調節遺伝子を提示している可能性がある（例えばＧａｔｔｉ，Ｒ．Ａ．他（１９８８）Ｎａｔｕｒｅ ３３６：５７７－５８０を参照）。転座、反転などに起因する、健常者、保有者、罹病者の三者間における染色体位置の相違を検出する場合にも、本発明のヌクレオチド配列を用い得る。

【０２２５】

本発明の別の実施態様では、ＡＴＲＳ、その触媒作用断片或いは免疫原断片またはそのオリゴペプチドを、種々の任意の薬物スクリーニング技術における、化合物のライブラリのスクリーニングに用い得る。薬物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ができる。ＡＴＲＳと、試験する薬剤との結合による複合体の形成を測定し得る。

【０２２６】

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（例えばＧｅｙｓｅｎ他（１９８４）ＰＣＴ出願番号 ＷＯ８４／０３５６４を参照）。この方法においては、多数の様々な低分子の試験用化合物を固体基板上で合成する。試験用化合物は、ＡＴＲＳ、或いはその断片と反応してから洗浄される。次に、本技術分野で周知の方法で、結合したＡＴＲＳを検出する。精製したＡＴＲＳはまた、上記した薬物スクリーニング技術に用いるプレート上に直接コーティングできる。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。

【０２２７】

別の実施態様では、ＡＴＲＳと特異結合可能な中和抗体がＡＴＲＳとの結合について試験化合物と競合する、競合的薬物スクリーニングアッセイを用い得る。この方法では、抗体を用いて、１つ以上の抗原決定基をＡＴＲＳと共有する、どのペプチドの存在をも検出できる。

【０２２８】

更なる実施態様では、ＡＴＲＳをコードするヌクレオチド配列群を、将来に開発される分

子生物学技術であり、現在知られているヌクレオチド配列群の諸特性（限定はされないが、トリプレット遺伝暗号及び特異的な塩基対相互作用など）に依存するあらゆる新技術に用い得る。

【０２２９】

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

【０２３０】

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６０／２５５，９６３号は、言及することをもって本明細書の一部とする。

【０２３１】

（実施例）

１　ｃＤＮＡライブラリの作製

Ｉｎｃｙｔｅ ｃＤＮＡ群の由来は、ＬＩＦＥＳＥＱ ＧＯＬＤデータベース（Ｉｎｃｙｔｅ Ｇｅｎｏｍｉｃｓ，Ｐａｌｏ Ａｌｔｏ ＣＡ）に記載されたｃＤＮＡライブラリ群である。幾つかの組織はホモジナイズしてグアニジニウムイソチオシアネート溶液に溶解し、他の組織はホモジナイズしてフェノールにまたは変性剤群の好適な混合液に溶解した。混合液の１例であるＴＲＩＺＯＬ（Ｌｉｆｅ Ｔｅｃｈｎｏｌｏｇｉｅｓ）は、フェノールとグアニジンイソチオシアネートとの単相溶液である。結果として得た溶解物は、塩化セシウムのクッション液の上に重層して遠心分離するか、クロロフォルムで抽出した。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、あるいは別の慣例的方法で、溶解物からＲＮＡを沈殿させた。

【０２３２】

ＲＮＡの純度を高めるため、フェノールによるＲＮＡの抽出および沈殿を、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。

【０２３３】

場合によってはＳｔｒａｔａｇｅｎｅ社にＲＮＡを提供し、対応するｃＤＮＡライブラリ群をＳｔｒａｔａｇｅｎｅ社が作製した。そうでない場合は、ＵＮＩＺＡＰベクターシステム（Ｓｔｒａｔａｇｅｎｅ）またはＳＵＰＥＲＳＣＲＩＰＴプラスミドシステム（Ｌｉｆｅ Ｔｅｃｈｎｏｌｏｇｉｅｓ）を用いて本技術分野で既知の推奨方法または類似の方法でｃＤＮＡを合成し、ｃＤＮＡライブラリを作製した（前出のＡｕｓｕｂｅｌ，１９９７，５．１－６．６ユニットなどを参照）。逆転写は、オリゴｄ（Ｔ）またはランダムプライマーを用いて開始した。合成オリゴヌクレオチドアダプターを二本鎖ｃＤＮＡに連結反応させ、好適な制限酵素または酵素群でｃＤＮＡを消化した。殆どのライブラリに対して、ｃＤＮＡのサイズ選択（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）、ある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ライゲーションされた。好適なプラスミドは、例えばＰＢＬＵＥＳＣＲＩＰＴプラスミド（Ｓｔｒａｔａｇｅｎｅ）、ＰＳＰＯＲＴ１プラスミド（Ｌｉｆｅ Ｔｅｃｈｎｏｌｏｇｉｅｓ）ＰＣＤＮＡ２．１プラスミド（Ｉｎｖｉｔｒｏｇｅｎ，Ｃａｒｌｓｂａｄ ＣＡ）、ＰＢＫ－ＣＭＶプラスミド（Ｓｔｒａｔａｇｅｎｅ）、ＰＣＲ２－ＴＯＰＯＴＡプラスミド（Ｉｎｖｉｔｒｏｇｅｎ）、ＰＣＭＶ－ＩＣＩＳプラスミド（Ｓｔｒａｔａｇｅｎｅ）、ｐＩＧＥＮ（Ｉｎｃｙｔｅ Ｇｅｎｏｍｉｃｓ，Ｐａｌｏ Ａｌｔｏ ＣＡ）、ｐＲＡＲＥ（Ｉｎｃｙｔｅ Ｇｅｎｏｍｉｃｓ）、またはｐｌ

ＮＣＹ（Ｉｎｃｙｔｅ Ｇｅｎｏｍｉｃｓ）、またはこれらの誘導体である。組換えプラスミドは、Ｓｔｒａｔａｇｅｎｅ社のＸＬ１－Ｂｌｕｅ、ＸＬ１－ＢＩｕｅＭＲＦまたはＳＯＬＲ、あるいはＬｉｆｅ Ｔｅｃｈｎｏｌｏｇｉｅｓ社のＤＨ５α、ＤＨ１０ＢまたはＥｌｅｃｔｒｏＭＡＸ ＤＨ１０Ｂなど適格な大腸菌細胞に形質転換した。

【０２３４】

２　ｃＤＮＡクローンの単離

実施例１のようにして得たプラスミドの、宿主細胞からの回収は、ＵＮＩＺＡＰベクターシステム（Ｓｔｒａｔａｇｅｎｅ）を用いたｉｎ ｖｉｖｏ切除によって、あるいは細胞溶解によって行った。プラスミドを精製する方法は、ＭａｇｉｃまたはＷＩＺＡＲＤ Ｍｉｎｉｐｒｅｐｓ ＤＮＡ精製システム（Ｐｒｏｍｅｇａ）、ＡＧＴＣ Ｍｉｎｉｐｒｅｐ精製キット（Ｅｄｇｅ Ｂｉｏｓｙｓｔｅｍｓ，Ｇａｉｔｈｅｒｓｂｕｒｇ ＭＤ）、ＱＩＡＧＥＮ社のＱＩＡＷＥＬＬ ８ Ｐｌａｓｍｉｄ、ＱＩＡＷＥＬＬ ８ Ｐｌｕｓ Ｐｌａｓｍｉｄおよびｑｉａｗｅｌｌ ８ Ｕｌｔｒａ Ｐｌａｓｍｉｄ 精製システム、Ｒ．Ｅ．Ａ．Ｌ． Ｐｒｅｐ ９６プラスミド精製キットの中から少なくとも１つを用いた。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。

【０２３５】

別法では、高処理フォーマットで直接結合ＰＣＲ法を用い、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）。宿主細胞の溶解および熱サイクリング過程は、単一反応混合液中で行った。サンプルを加工し、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）およびＦＬＵＯＲＯＳＫＡＮＩＩ蛍光スキャナ（Ｌａｂｓｙｓｔｅｍｓ Ｏｙ，Ｈｅｌｓｉｎｋｉ，Ｆｉｎｌａｎｄ）を用いて蛍光分析的に定量した。

【０２３６】

３　シークエンシングおよび分析

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、標準的方法あるいは高処理装置、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には、また、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プロトコルと塩基対呼び出しソフトウェアとを用いるＡＢＩ ＰＲＩＳＭ ３７３または３７７シークエンシングシステム（Ａｐｐｌｉｅｄ Ｂｉｏｓｙｓｔｅｍｓ）か、あるいはその他の本技術分野で既知の配列解析システムを用いた。ｃＤＮＡ配列内のリーディングフレームは、標準的方法（前出のＡｕｓｕｂｅｌ，１９９７，７．７ユニットに概説）を用いて同定した。いくつかのｃＤＮＡ配列を選択し、実施例８に開示する技術で配列を伸長させた。

【０２３７】

Ｉｎｃｙｔｅ ｃＤＮＡに由来するポリヌクレオチド配列は、ベクター、リンカーおよびポリ（Ａ）配列を除去し、あいまいな塩基をマスクすることによって有効性を確認した。その際、ＢＬＡＳＴと、動的プログラミングと、隣接ジヌクレオチド頻度分析とに基づく、アルゴリズムとプログラムとを用いた。次に、Ｉｎｃｙｔｅ ｃＤＮＡ配列またはそれらの翻訳の問い合わせを、以下のデータベース群に対して行った。すなわち、選抜した公

共のデータベース群（例えばＧｅｎＢａｎｋの霊長類、げっ歯類、哺乳類、脊椎動物、真核生物のデータベースと、ＢＬＯＣＫＳ、ＰＲＩＮＴＳ、ＤＯＭＯ、ＰＲＯＤＯＭ）と、ヒト、ラット、マウス、線虫（Ｃａｅｎｏｒｈａｂｄｉｔｉｓ ｅｌｅｇａｎｓ）、出芽酵母（Ｓａｃｃｈａｒｏｍｙｃｅｓ ｃｅｒｅｖｉｓｉａｅ）、分裂酵母（Ｓｃｈｉｚｏｓａｃｃｈａｒｏｍｙｃｅｓ ｐｏｍｂｅ）および鵞口瘡カンジダ（Ｃａｎｄｉｄａ ａｌｂｉｃａｎｓ）からの配列群を持つＰＲＯＴＥＯＭＥデータベース群（Ｉｎｃｙｔｅ Ｇｅｎｏｍｉｃｓ，Ｐａｌｏ Ａｌｔｏ ＣＡ）、および、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群（ＰＦＡＭ等）である（ＨＭＭは、遺伝子ファミリーのコンセンサス１次構造を分析する確率的アプローチである。例えば、Ｅｄｄｙ，Ｓ．Ｒ．（１９９６）Ｃｕｆｆ．Ｏｐｉｎ．Ｓｔｒｕｃｔ．Ｂｉｏｌ．６：３６１－３６５等を参照）。問合せは、ＢＬＡＳＴ、ＦＡＳＴＡ、ＢＬＩＭＰＳ、およびＨＭＭＥＲに基づくプログラムを用いて行った。Ｉｎｃｙｔｅ ｃＤＮＡ配列を構築し、完全長のポリヌクレオチド配列を産出した。あるいは、ＧｅｎＢａｎｋ ｃＤＮＡ群、ＧｅｎＢａｎｋ ＥＳＴ群、スティッチされた配列群、ストレッチされた配列群、またはＧｅｎｓｃａｎ予測コード配列群（実施例４および５を参照）を用い、Ｉｎｃｙｔｅ ｃＤＮＡの集団を完全長まで伸長させた。ＰｈｒｅｄとＰｈｒａｐとＣｏｎｓｅｄとに基づくプログラムを用いて構築し、ＧｅｎＭａｒｋとＢＬＡＳＴとＦＡＳＴＡと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ついてスクリーニングした。完全長ポリヌクレオチド配列を翻訳し、対応する完全長ポリペプチド配列を誘導した。あるいは、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。完全長ポリペプチド配列群の続いての分析としての問い合わせを、ＧｅｎＢａｎｋタンパク質データベース群（ｇｅｎｐｅｐｔ）、ＳｗｉｓｓＰｒｏｔ、ＰＲＯＴＥＯＭＥデータベース群、ＢＬＯＣＫＳ、ＰＲＩＮＴＳ、ＤＯＭＯ、ＰＲＯＤＯＭおよびＰｒｏｓｉｔｅ等のデータベースや、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群に対し行った。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）およびＬＡＳＥＲＧＥＮＥソフトウェア（ＤＮＡＳＴＡＲ）を用いて分析した。ポリヌクレオチドとポリペプチドとの配列アラインメントは、ＭＥＧＡＬＩＧＮマルチシークエンスアラインメントプログラム（ＤＮＡＳＴＡＲ）に組み込まれたＣＬＵＳＴＡＬアルゴリズムが指定する、デフォルトパラメータを用いて作製する。これは、アラインメントした配列間の一致率も計算する。

【０２３８】

表７は、Ｉｎｃｙｔｅ ｃＤＮＡと完全長配列との分析と構築とに利用したツールとプログラムとアルゴリズムとの概略と、適用可能な説明、参照文献、閾値パラメータを示す。用いたツール、プログラムおよ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は好適な参照文献であり、全ての文献は引用を以って全文を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、確率値など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い）。完全長ポリヌクレオチド配列およびポリペプチド配列の構築及び分析のための上記のプログラムは、ＳＥＱ ＩＤ ＮＯ：２のポリヌクレオチド配列断片の同定にも利用した。ハイブリダイゼーション及び増幅技術に有用な約２０～約４０００ヌクレオチドの断片を表４の列２に示す。

【０２３９】

４　ゲノムＤＮＡからのコード配列の同定および編集

推定上のアミノアシルｔＲＮＡ合成酵素は、公共のゲノム配列データベース（例えば、ｇｂｐｒｉやｇｂｈｔｇ）に対してＧｅｎｓｃａｎ遺伝子同定プログラムを実行して先ず同定された。Ｇｅｎｓｃａｎは汎用遺伝子同定プログラムであり、様々な生物からのゲノムＤＮＡ配列を分析する（Ｂｕｒｇｅ，Ｃ．およびＳ．Ｋａｒｌｉｎ（１９９７）Ｊ．Ｍｏｌ．Ｂｉｏｌ．２６８：７８－９４、Ｂｕｒｇｅ，Ｃ．およびＳ．Ｋａｒｌｉｎ（１９９８）Ｃｕｒｒ．Ｏｐｉｎ．Ｓｔｒｕｃｔ．Ｂｉｏｌ．８：３４６－３５

４参照）。このプログラムは予測されたエキソン群を連結し、メチオニンから終止コドンに及ぶ、構築されたｃＤＮＡ配列を形成する。Ｇｅｎｓｃａｎの出力は、ポリヌクレオチドおよ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予測ｃＤＮＡ配列の内、どの配列がアミノアシルｔＲＮＡ合成酵素をコードするかを判定するために、コードされるポリペプチドを、ＰＦＡＭモデルに対してアミノアシルｔＲＮＡ合成酵素について問合せて分析した。また、アミノアシルｔＲＮＡ合成酵素としてアノテーションが付けられたＩｎｃｙｔｅのｃＤＮＡ配列に対する相同性を基に、可能性のあるアミノアシルｔＲＮＡ合成酵素が同定された。これら選択したＧｅｎｓｃａｎ予測配列を、次にＢＬＡＳＴ分析により、公共データベースｇｅｎｐｅｐｔおよびｇｂｐｒｉと比較した。必要であれば、ｇｅｎｐｅｐｔからのトップＢＬＡＳＴヒットと比較することによりＧｅｎｓｃａｎ予測配列を編集し、余分なまたは省略されたエキソンなど、Ｇｅｎｓｃａｎが予測した配列におけるエラーを補正した。ＢＬＡＳＴ分析はまた、Ｇｅｎｓｃａｎ予測配列の、いかなるＩｎｃｙｔｅ ｃＤＮＡまたは公共ｃＤＮＡカバレッジ（ｃｏｖｅｒａｇｅ）の発見にも用いられ、したがって転写の証拠を提供した。Ｉｎｃｙｔｅ ｃＤＮＡカバレッジが利用できた場合には、この情報を用いてＧｅｎｓｃａｎ予測配列を補正または確認した。完全長ポリヌクレオチド配列は、実施例３に記載した構築プロセスを用いて、Ｉｎｃｙｔｅ ｃＤＮＡ配列および／または公共ｃＤＮＡ配列でＧｅｎｓｃａｎ予測コード配列を構築して得た。あるいは、完全長ポリヌクレオチド配列は、編集後または非編集のＧｅｎｓｃａｎ予測コード配列に、完全に由来する。

【０２４０】

５　ゲノム配列データのｃＤＮＡ配列データとの統合

スティッチ配列（Ｓｔｉｃｈｅｄ Ｓｅｑｕｅｎｃｅ）

部分ｃＤＮＡ配列群を伸長させるため、実施例４に記載のＧｅｎｓｃａｎ遺伝子同定プログラムが予測したエキソン群を用いた。実施例３に記載したように構築した部分ｃＤＮＡ群を、ゲノムＤＮＡにマッピングし、また、関連するｃＤＮＡ群と、１つ以上のゲノム配列からＧｅｎｓｃａｎ予測されたエキソン群とを有するクラスター群に分解した。ｃＤＮＡとゲノムとの情報を統合すべく、グラフ理論および動的プログラミングに基づく或るアルゴリズムを用いて各クラスターを分析し、潜在的スプライス変異配列群を産生した。配列群を続いて確認、編集または伸長し、完全長配列を創出した。区間の全長が、２つ以上の配列に在るような配列区間群をクラスター内で同定し、そのように同定された区間群は、推移性により、等しいと考えた。例えば、或る区間が、１つのｃＤＮＡと２つのゲノム配列とに在る場合、３つの区間は全て等しいと考えた。このプロセスにより、無関係だが連続したゲノム配列群を、ｃＤＮＡ配列で結び合わせて架橋し得る。このようにして同定した区間を、それらの親配列（ｐａｒｅｎｔ ｓｅｑｕｅｎｃｅｓ）に沿って現われる順にスティッチアルゴリズムで「縫い合わせ」、可能な限り最長の配列と変異配列群とを産生した。１種類の親配列に沿って続く区間間の連鎖（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）より優先した。結果として得たスティッチ配列群を翻訳し、ＢＬＡＳＴ分析で公共データベースｇｅｎｐｅｐｔおよびｇｂｐｒｉと比較した。Ｇｅｎｓｃａｎが予測した不正確なエキソン群は、ｇｅｎｐｅｐｔからのトップＢＬＡＳＴヒットとの比較により補正した。必要な場合、追加ｃＤＮＡ配列群を用いるかゲノムＤＮＡの検査により、配列群を更に伸長させた。

【０２４１】

ストレッチ配列（Ｓｔｒｅｔｃｈｅｄ Ｓｅｑｕｅｎｃｅ）

部分ＤＮＡ配列群を、ＢＬＡＳＴ分析に基づく１アルゴリズムにより完全長まで伸長した。先ず、ＢＬＡＳＴプログラムを用いて、ＧｅｎＢａｎｋの霊長類、げっ歯類、哺乳類、脊椎動物および真核生物のデータベースなどの公共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生し、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対し、キメラタンパク質内では挿入または欠失が起こり得る。ＧｅｎＢａｎｋタンパク質相同体、キメラタンパク質またはその両方をプローブとして用い、公共のヒトゲノムデータベースから相同ゲノム配列を検索した。このようにして、部分的なＤＮＡ配列を、相同ゲノム配列の付加によりストレッチすなわち伸長した。結果として得られるストレッチ配列を検査し、完全遺伝子を含んでいるか否かを決定した。

【０２４２】

６　ＡＴＲＳ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：２を構築するために用いた配列を、ＢＬＡＳＴ他のＳｍｉｔｈ－Ｗａｔｅｒｍａｎアルゴリズムの実装を用いて、Ｉｎｃｙｔｅ　ＬＩＦＥＳＥＱデータベース及び公共ドメインデータベース群の配列と比較した。ＳＥＱ　ＩＤ　ＮＯ：２と一致するこれらのデータベースの配列を、Ｐｈｒａｐなどの構築アルゴリズムを使用して、連続的配列及びオーバーラップした配列のクラスターに構築した（表７）。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などの公的な情報源から入手可能な放射線ハイブリッドおよび遺伝地図データを用いて、いずれかのクラスター化された配列が既にマッピングされているかを判定した。マッピングされた配列が或るクラスターに含まれている場合、そのクラスターの全配列が、個々の配列番号と共に、地図上の位置に割り当てられた。

【０２４３】

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アームの末端に対して測定する（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およびコールドスポットに起因して広範囲に変化する）。ｃＭ距離は、各クラスタ内に配列が含まれる放射線ハイブリッドマーカー類に対して境界を提供するＧｅｎｅｔｈｏｎによってマッピングされた遺伝マーカー群に基づく。ＮＣＢＩ「ＧｅｎｅＭａｐ’９９」（ｈｔｔｐ：／／ｗｗｗ．ｎｃｂｉ．ｎｌｍ．ｎｉｈ．ｇｏｖ／ｇｅｎｅｍａｐ／）など一般個人が入手可能なヒトゲノム地図などの資源を用いて、既に同定された疾患遺伝子類が、上記の区間内若しくは近傍にマップされているかを判定できる。

【０２４４】

７　ポリヌクレオチド発現の分析

ノーザン分析は、遺伝子の転写物の存在を検出するために用いられる実験技術であり、特定の細胞種または組織からのＲＮＡが結合している膜への、標識されたヌクレオチド配列のハイブリダイゼーションに関与する（例えば前出のＳａｍｂｒｏｏｋ，　７章、同Ａｕｓｕｂｅｌ　（１９９５）　４章および１６章を参照）。

【０２４５】

ＢＬＡＳＴを応用した類似のコンピュータ技術を用いて、ＧｅｎＢａｎｋやＬＩＦＥＳＥＱ（Ｉｎｃｙｔｅ　Ｇｅｎｏｍｉｃｓ）などのｃＤＮＡデータベースにおいて、同一または関連分子を検索した。ノーザン分析は、いくつかの膜系ハイブリダイゼーションよりも非常に速い。更に、任意の特定の一致を厳密なあるいは相同的なものとして分類するか否かを決定するため、コンピュータ検索の感度を修正することができる。検索の基準は積スコアであり、次式で定義される。

【０２４６】

【数１】

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

【０２４７】

積スコアは、２つの配列間の類似度と、配列が一致する長さとの両方を考慮している。積スコアは、０～１００の正規化された値であり、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。ＢＬＡＳＴスコアを計算するには、或る高スコアリングセグメント対（ＨＳＰ）内の一致する各塩基に＋５のスコアを割り当て、各不一致塩基に－４を割り当てる。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る）。２以上のＨＳＰがある場合には、最高ＢＬＡＳＴスコアのセグメント対を用いて積スコアを計算する。積スコアは、断片的オーバーラップとＢＬＡＳＴアラインメントの質とのバランスを表す。例えば積スコア１００は、比較した２つの配列の短い方の長さ全体にわたって１００％一致する場合のみ得られる。積スコア７０は、一端が１００％一致し、７０％オーバーラップしているか、他端が８８％一致し、１００％オーバーラップしているかのいずれかの場合に得られる。積スコア５０は、一端が１００％一致し、５０％オーバーラップしているか、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。

【０２４８】

或いは、ＡＴＲＳをコードするポリヌクレオチド配列を、由来する組織源について分析する。例えば幾つかの完全長配列は、少なくとも一部は、オーバーラップするＩｎｃｙｔｅ ｃＤＮＡ配列群を用いて構築される（実施例３を参照）。各ｃＤＮＡ配列は、ヒト組織から作製されたｃＤＮＡライブラリに由来する。各ヒト組織は、以下の臓器／組織カテゴリーの１つに分類される。すなわち心血管系、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、血液および免疫系、肝、筋骨格系、神経系、膵臓、呼吸器系、感覚器、皮膚、顎口腔系、非分類性／混合性または尿路である。各カテゴリーのライブラリ数を数えて、全カテゴリーの総ライブラリ数で除する。同様に、各ヒト組織は、以下の疾患／条件カテゴリーすなわち癌、細胞株、発達、炎症、神経性、外傷、心血管、プール、その他の１つに分類される。各カテゴリーのライブラリ数を数えて、全カテゴリーの総ライブラリ数で除する。得られるパーセンテージは、ＡＴＲＳをコードするｃＤＮＡの組織特異的および疾患特異的な発現を反映する。ｃＤＮＡ配列およびｃＤＮＡライブラリ／組織の情報は、ＬＩＦＥＳＥＱ ＧＯＬＤ データベース（Ｉｎｃｙｔｅ Ｇｅｎｏｍｉｃｓ, Ｐａｌｏ Ａｌｔｏ ＣＡ）を参照されたい。

【０２４９】

８ ＡＴＲＳをコードするポリヌクレオチドの伸長

完全長のポリヌクレオチド配列はまた、完全長分子の適切な断片から設計したオリゴヌクレオチドプライマー群を用いて該断片を伸長させて産生した。或る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、別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。開始プライマー群の設計にはＯＬＩＧＯ ４．０６ソフトウェア（Ｎａｔｉｏｎａｌ Ｂｉｏｓｃｉｅｎｃｅｓ）或いは別の適切なプログラムを用い、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～約７２℃の温度で標的配列にアニーリングするようにした。ヘアピン構造およびプライマー－プライマー二量体を生ずるようなヌクレオチド群の伸長は、全て回避した。

【０２５０】

選択したヒトｃＤＮＡライブラリ群を用い、配列を伸長した。２段階以上の伸長が必要または望ましい場合には、付加的プライマーあるいはプライマーのネステッドセットを設計した。

【０２５１】

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を有し、また、２００ ｎｍｏｌの各プライマーを有する。また、Mg^{2+}と$(NH_4)_2SO_4$と２－メルカプトエタノールとを含有する反応バッファーと、Ｔａｑ ＤＮＡポリメラーゼ（Ａｍｅｒｓｈａｍ Ｐｈａｒｍａｃｉａ Ｂｉｏｔｅｃｈ）と、ＥＬＯＮＧＡＳＥ酵素（Ｌｉｆｅ Ｔｅｃｈｎｏｌｏｇｉｅｓ）と、Ｐｆｕ ＤＮＡポリメラーゼ（Ｓｔｒａｔａｇｅｎｅ）

とを含有する。プライマー対であるＰＣＩ ＡとＰＣＩ Ｂとに対し、以下のパラメータで増幅した。
ステップ１　　９４℃で３分間
ステップ２　　９４℃で１５秒間
ステップ３　　６０℃で１分間
ステップ４　　６８℃で２分間
ステップ５　　ステップ２、３、及び４を２０回繰り返す
ステップ６　　６８℃で５分間
ステップ７　　４℃で保存
別法では、プライマー対であるＴ７とＳＫ＋に対し、以下のパラメータで増幅した。
ステップ１　　９４℃で３分間
ステップ２　　９４℃で１５秒間
ステップ３　　５７℃で１分間
ステップ４　　６８℃で２分間
ステップ５　　ステップ２、３、及び４を２０回繰り返す
ステップ６　　６８℃で５分間
ステップ７　　４℃で保存
各ウェルのＤＮＡ濃度は、１× 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（Ｃｏｒｎｉｎｇ Ｃｏｓｔａｒ，Ａｃｔｏｎ ＭＡ）の各ウェルに分配してＤＮＡが試薬と結合できるようにして測定した。サンプルの蛍光を計測してＤＮＡの濃度を定量すべく、プレートをＦｌｕｏｒｏｓｋａｎ ＩＩ（Ｌａｂｓｙｓｔｅｍｓ Ｏｙ，Ｈｅｌｓｉｎｋｉ，Ｆｉｎｌａｎｄ）でスキャンした。反応混合物のアリコット５～１０μｌを１％アガロース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判定した。

【０２５２】

伸長したヌクレオチドは、脱塩および濃縮して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、音波処理またはせん断し、ｐＵＣ １８ベクター（Ａｍｅｒｓｈａｍ Ｐｈａｒｍａｃｉａ Ｂｉｏｔｅｃｈ）への再連結を行った。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。伸長させたクローンをＴ４リガーゼ（Ｎｅｗ Ｅｎｇｌａｎｄ Ｂｉｏｌａｂｓ，Ｂｅｖｅｒｌｙ ＭＡ）を用いてｐＵＣ １８ベクター（Ａｍｅｒｓｈａｍ Ｐｈａｒｍａｃｉａ Ｂｉｏｔｅｃｈ）に再連結し、Ｐｆｕ ＤＮＡポリメラーゼ（Ｓｔｒａｔａｇｅｎｅ）で処理して制限部位のオーバーハングを満たし、適格な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有する培地に移し、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穴プレートに３７℃で一晩培養した。

【０２５３】

細胞を溶解し、Ｔａｑ ＤＮＡポリメラーゼ（Ａｍｅｒｓｈａｍ Ｐｈａｒｍａｃｉａ Ｂｉｏｔｅｃｈ）およびＰｆｕ ＤＮＡポリメラーゼ（Ｓｔｒａｔａｇｅｎｅ）を用いて以下の手順でＤＮＡをＰＣＲ増幅した。
ステップ１　　９４℃で３分間
ステップ２　　９４℃で１５秒間
ステップ３　　６０℃で１分間
ステップ４　　７２℃で２分間
ステップ５　　ステップ２、３、及び４を２９回繰り返す
ステップ６　　７２℃で５分間

ステップ７　　４℃で保存
ＤＮＡは、上記のＰＩＣＯＧＲＥＥＮ試薬（Ｍｏｌｅｃｕｌａｒ Ｐｒｏｂｅｓ）によって定量した。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、およびＤＹＥＮＡＭＩＣ ＤＩＲＥＣＴ ｋｉｔ（Ａｍｅｒｓｈａｍ Ｐｈａｒｍａｃｉａ Ｂｉｏｔｅｃｈ）またはＡＢＩ ＰＲＩＳＭ ＢＩＧＤＹＥ ターミネーターサイクル シークエンシングレディ反応キット（Ｔｅｒｍｉｎａｔｏｒ ｃｙｃｌｅ ｓｅｑｕｅｎｃｉｎｇ ｒｅａｄｙ ｒｅａｃｔｉｏｎ ｋｉｔ）（Ａｐｐｌｉｅｄ Ｂｉｏｓｙｓｔｅｍｓ）を用いてシークエンシングした。

【０２５４】
同様に、上記手順を用いて完全長ポリヌクレオチド配列を検証した。あるいは、完全長ポリヌクレオチドを用い、上記手順で、そのような伸長のために設計したオリゴヌクレオチド類と、或る適切なゲノムライブラリとを用いて５’調節配列を得た。

【０２５５】
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：２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本質的に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 ビーズカラム（Ａｍｅｒｓｈａｍ Ｐｈａｒｍａｃｉａ Ｂｉｏｔｅｃｈ）を用いて実質的に精製する。Ａｓｅ Ｉ、Ｂｇｌ ＩＩ、Ｅｃｏ ＲＩ、Ｐｓｔ Ｉ、ＸｂａＩまたはＰｖｕ ＩＩ（ＤｕＰｏｎｔ ＮＥＮ）のいずれか１つのエンドヌクレアーゼで消化されたヒトゲノムＤＮＡの、典型的な膜ベースのハイブリダイゼーション解析において、毎分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。

【０２５６】
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。非特異的シグナル群を除去するため、最大で例えば０．１×クエン酸ナトリウム食塩水および０．５％ドデシル硫酸ナトリウムの条件下で、ブロット群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。

【０２５７】
１０　マイクロアレイ
或るマイクロアレイ上でのアレイエレメント群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などを参照）、機械的マイクロスポッティング技術およびこれらから派生した技術を用いて達成し得る。上記各技術において基板は、均一な、非多孔性の表面を持つ固体とすべきである（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、ガラスチップおよびシリコンウエハがある。あるいは、ドットブロット法またはスロットブロット法に類似した手順を利用し、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および結合させてもよい。通常のアレイは、利用可能な、当業者に公知の方法と機械とを用いて作製でき、任意の適正数のエレメントを有し得る（例えばＳｃｈｅｎａ，Ｍ．他（１９９５）Ｓｃｉｅｎｃｅ ２７０：４６７－４７０

、Ｓｈａｌｏｎ．Ｄ．他（１９９６）Ｇｅｎｏｍｅ Ｒｅｓ．６：６３９－６４５、Ｍａｒｓｈａｌｌ，Ａ．およびＪ．Ｈｏｄｇｓｏｎ（１９９８）Ｎａｔ．Ｂｉｏｔｅｃｈｎｏｌ．１６：２７－３１を参照）。

【０２５８】

完全長ｃＤＮＡ、発現配列タグ（ＥＳＴ）、またはそれらの断片またはオリゴマーが、マイクロアレイのエレメントと成り得る。ハイブリダイゼーションに好適な断片またはオリゴマーを、ＬＡＳＥＲＧＥＮＥソフトウェア（ＤＮＡＳＴＡＲ）など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。アレイエレメント群を、生体サンプル中のポリヌクレオチド群とハイブリダイズする。生体サンプル中のポリヌクレオチドは、検出を容易にするために蛍光標識などの分子タグに抱合させる。ハイブリダイゼーション後、生体サンプルからのハイブリダイズされていないヌクレオチドを除去し、蛍光スキャナを用いて各アレイエレメントでのハイブリダイゼーションを検出する。あるいは、レーザ脱離およ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或る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、相補性の度合と相対存在度とを算定し得る。 一実施態様におけるマイクロアレイの調整および使用について、以下に詳述する。

【０２５９】

組織または細胞サンプルの調製

全ＲＮＡを組織サンプル群から単離するためグアニジニウム チオシアネート法を用い、ポリ（Ａ）$^{+}$ＲＮＡを精製するためオリゴ（ｄＴ）セルロース法を用いる。各ポリ（Ａ）$^{+}$ＲＮＡサンプルを、ＭＭＬＶ逆転写酵素、０．０５ｐｇ／μｌのオリゴ（ｄＴ）プライマー（２１ｍｅｒ）、１×第一鎖バッファー、０．０３ｕｎｉｔ／μｌのＲＮアーゼ阻害因子、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、２００ｎｇのポリ（Ａ）$^{+}$ＲＮＡを含有する体積２５ｍｌで行う。特異的対照ポリ（Ａ）$^{+}$ＲＮＡ群は、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。サンプル群の精製には、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）を用いる。混合後、２つの反応サンプルのエタノール沈殿を、１ｍｌのグリコーゲン（１ｍｇ／ｍｌ）、６０ｍｌの酢酸ナトリウム、および３００ｍｌの１００％エタノールで行う。サンプルは次に、ＳｐｅｅｄＶＡＣ（Ｓａｖａｎｔ Ｉｎｓｔｒｕｍｅｎｔｓ Ｉｎｃ．，Ｈｏｌｂｒｏｏｋ ＮＹ）を用いて完全に乾燥させ、１４μｌの５×ＳＳＣ／０．２％ＳＤＳ中で再懸濁する。

【０２６０】

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作製する。各アレイエレメントは、クローン化したｃＤＮＡインサート群により、ベクター含有細菌細胞群から増幅する。ＰＣＲ増幅は、ｃＤＮＡインサートの側面に位置するベクター配列群に相補的なプライマー類を用いる。３０サイクルのＰＣＲによって、１～２ｎｇの初期量から５μｇを超える最終量まで、アレイエレメントを増幅する。増幅したアレイエレメントは、ＳＥＰＨＡＣＲＹＬ－４００（Ａｍｅｒｓｈａｍ Ｐｈａｒｍａｃｉａ Ｂｉｏｔｅｃｈ）を用いて精製する。

【０２６１】

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、０．１％のＳＤＳおよびアセトン中で超音波洗浄し、処理中および処理後に多量の蒸留水で洗浄する。スライドグラスは、４％フッ化水素酸（ＶＷＲ Ｓｃｉｅｎｔｉｆｉｃ Ｐｒｏｄｕｃｔｓ Ｃｏｒｐｏｒａｔｉｏｎ（ＶＷ

Ｒ），Ｗｅｓｔ Ｃｈｅｓｔｅｒ ＰＡ）中でエッチングし、多量の蒸留水中で洗浄し、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は、１１０℃のオーブンで硬化させる。

【０２６２】

コーティングしたガラス基板にアレイエレメント群を付加する手順は、米国特許第５，８０７，５２２号に記載されている。この特許は、引用を以って本明細書の一部とする。平均濃度１００ｎｇ／μｌのアレイエレメントＤＮＡ１μｌを、高速ロボット装置（ｒｏｂｏｔｉｃ ａｐｐａｒａｔｕｓ）により、開放型キャピラリープリンティングエレメント（ｏｐｅｎ ｃａｐｉｌｌａｒｙ ｐｒｉｎｔｉｎｇ ｅｌｅｍｅｎｔ）に充填する。装置はここで、スライド毎に約５ｎｌのアレイエレメントサンプルを置く。

【０２６３】

マイクロアレイをＵＶ架橋するため、ＳＴＲＡＴＡＬＩＮＫＥＲ ＵＶ架橋剤（Ｓｔｒａｔａｇｅｎｅ）を用いる。マイクロアレイの洗浄を、室温において、０．２％ＳＤＳで１回、蒸留水で３回行う。非特異結合部位をブロックするため、マイクロアレイのインキュベートを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分間行った後、前に行ったように０．２％ＳＤＳおよび蒸留水で洗浄する。

【０２６４】

ハイブリダイゼーション

ハイブリダイゼーション反応に用いる９μｌのサンプル混合体には、Ｃｙ３またはＣｙ５で標識したｃＤＮＡ合成産物群の各０．２μｇを、５×ＳＳＣ，０．２％ＳＤＳハイブリダイゼーション緩衝液中に含む。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。アレイを、顕微鏡用スライドよりわずかに大きい空洞を有する防水チェンバーに移す。チェンバー内を湿度１００に保持するため、１４０μｌの５×ＳＳＣをチェンバーの１コーナーに加える。アレイを入れたチェンバーは、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分間ずつ３回洗浄し、乾燥させる。

【０２６５】

検出

レポーター標識したハイブリダイゼーション複合体を検出するには、Ｃｙ３の励起のために４８８ｎｍ、Ｃｙ５の励起のために６３２ｎｍでスペクトル線を発生し得るＩｎｎｏｖａ７０混合ガス１０Ｗレーザ（Ｃｏｈｅｒｅｎｔ，Ｉｎｃ．，Ｓａｎｔａ ＣｌａｒａＣＡ）を備えた顕微鏡を用いる。励起レーザ光の焦点をアレイ上に置くため、２０×顕微鏡対物レンズ（Ｎｉｋｏｎ，Ｉｎｃ．，Ｍｅｌｖｉｌｌｅ ＮＹ）を用いる。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、対物レンズを通してラスタースキャンする。本実施例で用いる１．８ｃｍ×１．８ｃｍのアレイは、解像度２０μｍでスキャンする。

【０２６６】

異なる２回のスキャンで、混合ガスマルチラインレーザは２つの蛍光色素を連続的に励起する。発光された光は、波長に基づき分離され、２つの蛍光色素に対応する２つの光電子増倍管検出器（ＰＭＴ Ｒ１４７７，Ｈａｍａｍａｔｓｕ Ｐｈｏｔｏｎｉｃｓ Ｓｙｓｔｅｍｓ，Ｂｒｉｄｇｅｗａｔｅｒ ＮＪ）に送られる。好適なフィルター群をアレイと光電子増倍管との間に設置して、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。

【０２６７】

通常、各スキャンの感度を較正するため、ｃＤＮＡ対照種を或る既知濃度でサンプル混合体に添加し、対照種が発生するシグナル強度を用いる。アレイ上の或る特定の位置には或

る相補的ＤＮＡ配列が含まれ、その位置におけるシグナルの強度をハイブリダイゼーション種の重量比１：１００，０００で相関させる。異なる源泉（例えば代表的な試験細胞および対照細胞など）からの２つのサンプルを、各々異なる蛍光色素で標識し、異なる発現をする遺伝子群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、較正するｃＤＮＡのサンプルを２種の蛍光色素で標識し、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。

【０２６８】

光電子増倍管の出力をディジタル化するため、ＩＢＭコンパチブルＰＣコンピュータにインストールした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）を用いる。ディジタル化したデータは、青色（低シグナル）から赤色（高シグナル）までの擬似カラースケールへのリニア２０色変換を用いて、シグナル強度がマッピングされた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および測定する場合には、各蛍光体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）を補正される。

【０２６９】

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。

【０２７０】

１１　相補的ポリヌクレオチド

ＡＴＲＳ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ＡＴＲＳの発現を検出、低減、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、これより小さなあるいは大きな配列の断片の場合でも、本質的に同じ手順を用いる。Ｏｌｉｇｏ４．０６ソフトウェア（Ｎａｔｉｏｎａｌ Ｂｉｏｓｃｉｅｎｃｅｓ）及びＡＴＲＳ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防止する。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、ＡＴＲＳをコードする転写物にリボソームが結合するのを阻害する。

【０２７１】

１２　ＡＴＲＳの発現

ＡＴＲＳの発現及び精製を、細菌若しくはウイルスを基にした発現系を用いて行う。細菌内でＡＴＲＳを発現させるには、抗生物質耐性遺伝子と、ｃＤＮＡ転写レベルを高める誘導性プロモーターとを有する好適なベクターにｃＤＮＡをサブクローニングする。このようなプロモーターとしては、ｌａｃオペレーター調節エレメントと併用するＴ５またはＴ７バクテリオファージプロモーター、およびｔｒｐ－ｌａｃ（ｔａｃ）ハイブリッドプロモーターが含まれるが、これらに限定するものではない。組換えベクターを、ＢＬ２１（ＤＥ３）などの好適な細菌宿主に形質転換する。抗生物質耐性細菌は、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ＡＴＲＳを発現する。真核細胞でのＡＴＲＳの発現は、昆虫細胞株または哺乳類細胞株に、一般にバキュロウイルスとして知られている*Ａｕｔｏｇｒａｐｈｉｃａ ｃａｌｉｆｏｒｎｉｃａ*核多角体病ウイルス（ＡｃＭＮＰＶ）の組換え型を感染させて行う。バキュロウイルスの非必須の多角体遺伝子を、相同組換えによって、或いはトランスファープラスミドの媒介を伴う細菌の媒介による遺伝子転移によって、ＡＴＲＳをコードするｃＤＮＡと置換する。ウイルスの感染力は維持され、強力な多角体プロモーターによって高レベルのｃＤＮＡ転写が行われる。組換えバキュロウイルスは、多くの場合は夜蛾の１種*Ｓｐｏｄｏｐｔｅｒａ ｆｒｕｇｉｐｅｒｄａ*（Ｓｆ９）昆虫細胞への感染に用いるが、ヒト肝細胞への感染に用いることもある。後者の感染の場合は、バキュロウイルスへの更なる遺伝的修飾が必要になる（Ｅｎｇｅ

lhard. E. K.他（1994）Proc. Natl. Acad. Sci. USA 91：3224－3227、Sandig, V. 他（1996）Hum. Gene Ther. 7：1937－1945を参照）。

【0272】

殆どの発現系では、ATRSが、例えばグルタチオンSトランスフェラーゼ（GST）と、またはFLAGや6－Hisなどのペプチドエピトープ標識と合成された融合タンパク質となるため、未精製の細胞溶解物からの組換え融合タンパク質の親和性ベースの精製を迅速に1ステップで行い得る。GSTは日本住血吸虫からの26kDaの酵素であり、タンパク質の活性および抗原性を維持した状態で、固定化したグルタチオン上での融合タンパク質の精製を可能とする（Amersham Pharmacia Biotech）。精製後、GST部分を、特異的に操作した部位でATRS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およびポリクローナル抗FLAG抗体（Eastman Kodak）を用いた免疫親和性精製を可能にする。6ヒスチジン残基が連続して伸長した6－Hisは、金属キレート樹脂上での精製を可能にする（QIAGEN）。タンパク質の発現および精製の方法は、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ATRSを直接用いて以下の実施例16、17、及び18の、適用可能なアッセイを行い得る。

【0273】

13 機能的アッセイ

ATRS機能を、哺乳類細胞培養系において生理学的に高められたレベルでの、ATRSをコードする配列の発現によって算定する。cDNAを高いレベルで発現する強いプロモーターを持つ哺乳類発現ベクターにcDNAをサブクローニングする。選択されるベクターとしては、PCMV SPORT（Life Technologies）およびPCR 3.1（Invitrogen, Carlsbad CA）があり、どちらもサイトメガロウイルスプロモーターを持つ。リポソーム製剤ある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、組換えベクターからのcDNA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、それら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ある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、前方散乱光と90°側方散乱光によって計測される細胞サイズと顆粒性の変化、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および細胞内におけるタンパク質の発現の変容、およ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容とがある。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。

【0274】

遺伝子発現におけるATRSの影響は、ATRSをコードする配列とCD64またはCD64－GFPのどちらかとが形質移入された、高度に精製された細胞集団を用いて算定できる。CD64またはCD64－GFPは、形質転換された細胞表面で発現し、ヒト免疫グロブリンG（IgG）の保存された複数の領域に結合する。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

ビーズを用いて効果的に分離される（ＤＹＮＡＬ．　Ｌａｋｅ　Ｓｕｃｃｅｓｓ．　ＮＹ）。ｍＲＮＡは、当業者に周知の方法で細胞から精製できる。ＡＴＲＳと目的の他の遺伝子とをコードするｍＲＮＡの発現は、ノーザン分析やマイクロアレイ技術で分析できる。

【０２７５】

１４　ＡＴＲＳに特異的な抗体の作製

ポリアクリルアミドゲル電気泳動法（ＰＡＧＥ；例えば、Ｈａｒｒｉｎｇｔｏｎ，　Ｍ．Ｇ．（１９９０）　Ｍｅｔｈｏｄｓ　Ｅｎｚｙｍｏｌ．　１８２：４８８－４９５を参照）または他の精製技術で実質的に精製されたＡＴＲＳを用いて、標準的なプロトコルでウサギを免疫化して抗体を作り出す。

【０２７６】

或いは、ＬＡＳＥＲＧＥＮＥソフトウェア（ＤＮＡＳＴＡＲ）を用いてＡＴＲＳアミノ酸配列を解析し、免疫原性の高い領域を判定する。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既知の種々の方法で抗体を生成する。例えばＣ末端付近或いは親水性領域内等の、適切なエピトープの選択方法については、当分野に記述が多い（前出のＡｕｓｕｂｅｌ，　１９９５，　１１章等を参照）。

【０２７７】

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（例えば前出のＡｕｓｕｂｅｌ，　１９９５を参照）。完全フロイントアジュバントにおいて、オリゴペプチド－ＫＬＨ複合体を用いてウサギを免疫化する。得られた抗血清の抗ペプチド活性及び抗ＡＴＲＳ活性を試験するには例えば、ペプチドまたはＡＴＲＳを基板に結合し、１％ＢＳＡを用いてブロッキング処理し、ウサギ抗血清と反応させて洗浄し、更に、放射性ヨウ素標識されたヤギ抗ウサギＩｇＧと反応させる。

【０２７８】

１５　特異的抗体を用いる天然ＡＴＲＳの精製

天然ＡＴＲＳ或いは組換えＡＴＲＳを、ＡＴＲＳに特異的な抗体を用いるイムノアフィニティークロマトグラフィにより実質的に精製する。イムノアフィニティーカラムは、ＣＮＢｒ－活性化したＳＥＰＨＡＲＯＳＥ（Ａｍｅｒｓｈａｍ　Ｐｈａｒｍａｃｉａ　Ｂｉｏｔｅｃｈ）など活性化クロマトグラフィー用レジンと抗ＡＴＲＳ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このレジンをブロックし、洗浄する。

【０２７９】

ＡＴＲＳを有する培養液をイムノアフィニティーカラムに通し、ＡＴＲＳ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バッファーで）そのカラムを洗浄する。そのカラムを、抗体とＡＴＲＳとの結合を切るような条件で（例えば、ｐＨ２～３のバッファー、或いは高濃度の尿素またはチオシアン酸イオンのようなカオトロープで）溶出させ、ＡＴＲＳを収集する。

【０２８０】

１６　ＡＴＲＳと相互作用する分子の同定

ＡＴＲＳまたは、生物学的に活性なその断片を、^{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分子を、標識したＡＴＲＳと共にインキュベートし、洗浄して、標識されたＡＴＲＳ複合体を有する全てのウェルをアッセイする。様々なＡＴＲＳ濃度で得られたデータを用いて、候補分子と結合したＡＴＲＳの数量及び親和性、会合についての値を計算する。

【０２８１】

別法では、ＡＴＲＳと相互作用する分子を、Ｆｉｅｌｄｓ，Ｓ．及びＯ．Ｓｏｎｇ（１９８９，Ｎａｔｕｒｅ３４０：２４５－２４６）に記載の酵母２－ハイブリッドシステム（ｙｅａｓｔ ｔｗｏ－ｈｙｂｒｉｄ ｓｙｓｔｅｍ）や、ＭＡＴＣＨＭＡＫＥＲシステム（Ｃｌｏｎｔｅｃｈ）などの２－ハイブリッドシステムに基づいた市販のキットを用いて分析する。

【０２８２】

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

【０２８３】

１７　ＡＴＲＳ活性の実証

ｔＲＮＡ合成酵素活性は、［^{14}Ｃ］－標識されたアミノ酸の存在下で、或る基質ｔＲＮＡのアミノアシル化として測定される。ＡＴＲＳは緩衝液中で、［^{14}Ｃ］－標識されたアミノ酸と、好適な同族ｔＲＮＡと共にインキュベートされる（例えば［^{14}Ｃ］バリンおよびｔＲＮＡ^{val}）。^{14}Ｃ－標識の産物は遊離［^{14}Ｃ］アミノ酸からクロマトグラフィーによって分離され、取り込まれた［^{14}Ｃ］はシンチレーションカウンタで定量される。このアッセイでは検出した^{14}Ｃ－標識産物の量が、ＡＴＲＳの活性に比例する。

【０２８４】

１８　ＡＴＲＳアゴニスト及びアンタゴニストの同定

ＡＴＲＳ活性化若しくは抑制のアゴニスト若しくはアンタゴニストを、実施例１７で記述したアッセイを用いてテストし得る。アゴニストはＡＴＲＳ活性の増加を生じ、アンタゴニストはＡＴＲＳ活性の低下を生じる。

【０２８５】

当業者には、本発明の要旨および精神から逸脱しない範囲での、記載した本発明の方法およびシステムの、種々の修正および変更の手段は自明であろう。本発明について説明するにあたり幾つかの実施例に関連して説明を行ったが、本発明の請求の範囲が、そのような特定の実施例に不当に制限されるべきではないことを理解されたい。分子生物学または関連分野の専門家には明らかな、本明細書に記載する本発明の実施方法の様々な修正は、明確に特許請求の範囲内にあるものとする。

【０２８６】

（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およびポリペプチド配列の命名法の概略を示す。

【０２８７】

表２は、ＧｅｎＢａｎｋ識別番号と、本発明のポリペプチドに最も近いＧｅｎＢａｎｋ相同体の注釈とを示す。また、各ポリペプチドとその相同体（１つ以上）が一致する確率スコアも併せて示す。

【０２８８】

表３は、予測されるモチーフおよびドメインなど、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方法、アルゴリズムおよび検索可能なデータベースと共に示す。

【０２８９】

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断片やゲノム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。

【０２９０】

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。

【０２９１】

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

【０２９２】

表７は、本発明のポリヌクレオチドとポリペプチドの分析に用いたツール、プログラム、アルゴリズムを、適用可能な説明、参照文献およ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
60211391	1	60211391CD1	2	60211391CB1

【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	6021139ICD1	g4590328	3.5e-271	バリル-tRNA 合成酵素 [マウス] Snoek, M. および van Vugt, H. (1999) The sequence and organization of the mouse valyl-tRNA synthetase gene G7a/Bat6 located in the MHC class III region. Immunogenetics 49:468-470

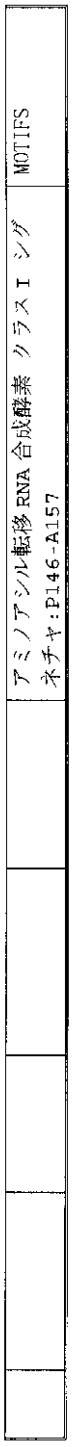
【表３】

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	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	60211391CD1	1063	S8 S40 S68 S72 S105 S288 S398 S447 S555 S613 S660 S674 S735 S741 S776 S832 S844 S904 S1019 S1039 T30 T336 T337 T567 T617 T1014 T1017 T1025 Y84 Y114 Y474	N147 N278	膜貫通ドメイン：S613-T634， V841-P858 （N 末端はサイトゾル内）	TMAP
					tRNA 合成酵素 クラス I（I、L、M および V）： A112-E794	HMMER-PFAM
					アミノアシル転移 RNA 合成酵素 クラス I シグネチャ BL00178： V148-A157，Q655-N665	BLIMPS-BLOCKS
					アミノアシル転移 RNA 合成酵素 クラス I シグネチャ：P132-P181	PROFILESCAN
					バリル tRNA 合成酵素 シグネチャ PR00986： F139-G150，H351-P368， R467-W480，D579-P600，Y610-R628	BLIMPS-PRINTS
					合成酵素 アミノアシル TRNA リガーゼ 生合成 ATP 結合 ロイシル TRNA バリル TRNA LEURS ロイシン TRNA PD000389： W113-V393，W529-Y610，I416-W586	BLAST-PRODOM
					合成酵素 アミノアシル TRNA バリル TRNA バリル TRNA バリン TRNA リガーゼ VALRS 生合成 PD118286：Y104-E293	BLAST-PRODOM
					合成酵素 バリル TRNA アミノアシル TRNA バリン TRNA リガーゼ VALRS 生合成 ATP 結合 896AA PD000939：V391-I533	BLAST-PRODOM
					合成酵素 アミノアシル TRNA リガーゼ 生合成 ATP 結合 バリル TRNA バリン TRNA VALRS イソロイシル TRNA PD000476：P593-R926，P146-L166	BLAST-PRODOM
					アミノアシル転移 RNA 合成酵素 クラス I DM00514｜P26640｜506-1094：V312-P886 DM00514｜P49696｜455-1047：V312-A888 DM00514｜P07806｜352-933：V312-P886 DM00514｜P28350｜341-921：V312-S887	BLAST-DOMO

【表４】

表3－2



【表５】

表4

ポリヌクレオチド SEQ ID NO:/ Incyte ID/ 配列長	配列断片
2/ 6021139lCB1/ 3557	1-446, 1-678, 1-685, 1-755, 12-825, 22-649, 24-825, 29-812, 43-827, 46-297, 64-680, 65-829, 70-827, 79-770, 87-345, 89-827, 96-512, 96-552, 96-827, 97-343, 97-429, 106-299, 158-823, 192-463, 697-1299, 746-1286, 746-1367, 804-1117, 823-1245, 823-1300, 823-1348, 823-1358, 849-1422, 1045-1715, 1092-1647, 1109-1447, 1109-1561, 1109-1594, 1109-1680, 1109-1752, 1109-1808, 1109-1838, 1109-1846, 1109-1920, 1115-1562, 1156-1753, 1169-1765, 1178-1774, 1178-1979, 1199-1845, 1201-1963, 1215-1710, 1217-1883, 1234-1878, 1260-1835, 1269-1734, 1279-1635, 1282-1825, 1290-2088, 1323-1885, 1327-1689, 1327-1870, 1339-1932, 1398-2046, 1492-2402, 1557-1600, 1557-2363, 1574-2374, 1577-2287, 1578-2311, 1599-2346, 1629-2378, 1638-2396, 1650-2475, 1694-2254, 1711-2400, 1721-2371, 1728-2304, 1756-2465, 1761-2359, 1763-2404, 1771-2324, 1773-2383, 1793-2417, 1805-2468, 1824-2459, 1834-2304, 1857-2557, 1863-2396, 1866-2596, 1872-2683, 1882-2679, 1891-2380, 1897-2528, 1901-2682, 1908-2430, 1908-2580, 1911-2387, 1911-2518, 1924-2556, 1926-2466, 1932-2700, 1989-2732, 1993-2695, 2019-2679, 2061-2481, 2061-2482, 2061-2534, 2061-2587, 2065-2711, 2095-2703, 2096-2705, 2098-2792, 2107-2595, 2111-2733, 2122-2738, 2147-2703, 2152-2733, 2173-2726, 2181-2591, 2184-2801, 2185-2747, 2193-3051, 2201-2711, 2207-2726, 2226-2747, 2234-2685, 2236-2809, 2240-2702, 2263-2736, 2283-2843, 2287-2721, 2288-2876, 2291-2737, 2297-2862, 2300-2736, 2309-3003, 2334-3160, 2350-2736, 2351-2878, 2354-2721, 2355-2900, 2372-3029, 2390-2928, 2393-2899, 2410-2919, 2429-2916, 2430-3265, 2442-3278, 2455-2904, 2459-3172, 2469-2801, 2474-3120, 2478-3034, 2490-3203, 2491-3146, 2517-3328, 2523-2982, 2526-3111, 2537-3102, 2565-3099, 2573-3227, 2582-3185, 2591-3271, 2597-3223, 2629-3262, 2635-3256, 2655-3218, 2669-3229, 2671-3329, 2692-3555, 2754-3377, 2762-3249, 2763-3252, 2766-3319, 2782-3482, 2783-3451, 2793-3501, 2827-3556, 2878-3407, 2885-3556, 2888-3406, 2892-3451, 2900-3385, 2900-3392, 2900-3397, 2911-3397, 2933-3556, 2953-3296, 2953-3556, 2967-3557, 2989-3489, 3001-3451, 3100-3220

【表６】

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
2	60211391CB1	BRAITDR03

【表７】

表6

ライブラリ	ベクター	ライブラリの説明
BRAITDR03	PCDNA2.1	このランダムにプライムされたライブラリは、胆管癌で死亡した55才白人女性から切除した異種皮質（不等皮質）（帯状回後部組織）から単離したRNAを用いて作製した。病理は主に弓隆部全体に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、内嗅皮質および中脳水道周囲灰白質領域にいくぶん神経原繊維濃縮体の散在があった。関連腫瘍組織の病理は、残存腫瘍または再発性腫瘍のある向分化型の肝臓胆管癌を示した。患者の病歴は胆管癌、術後バッド・キアーリ症候群、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。

【表８】

表7－1

プログラム	説明	参照文献	パラメータ閾値
ABI FACTURA	ベクター配列を除き、核酸配列のまぎらわしい塩基をマスクするプログラム。	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	アミノ酸または核酸配列の比較、注釈に有用なファーストデータファインダ。	Applied Biosystems, Foster City, CA. Paracel Inc., Pasadena, CA.	不一致<50%
ABI AutoAssembler	核酸配列の構築を行うプログラム。	Applied Biosystems, Foster City, CA.	
BLAST	アミノ酸および核酸配列の配列類似性検索に有用なベーシックローカルアラインメント検索ツール（Basic Local Alignment Search Tool）。BLASTには5つの機能がある：blastp、blastn、blastx、tblastnおよびtblastx。	Altschul, S.F. 他 (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. 他 (1997) Nucleic Acids Res. 25:3389-3402.	EST：確率値=1.0E-8 以下 完全長配列：確率値=1.0E-10以下
FASTA	一群の同じタイプの配列と問合せ配列との間の類似性について検索するPearsonおよびLipmanアルゴリズム。FASTAは最低5つの機能からなる：fasta、tfasta、fastx、tfastxおよびssearch。	Pearson, W.R. およびD.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98, Smith, T.F. およびM.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	EST：fasta E 値=1.06E-6 構築された EST:fasta 同一性=95%以上 一致した長さ=200塩基以上 fasta E 値=1.0E-8以下 完全長配列： fastx スコア=100以上
BLIMPS	配列をBLOCKS、PRINTS、DOMO、PRODOMおよびPFAM データベースの配列と対応させて遺伝子ファミリ、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 Searcher。	Henikoff, S. およびJ.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. およびS. Henikoff (1996) Methods Enzymol. 266:88-105、Attwood, T.K. 他. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	確率値= 1.0E-3以下
IMMER	問合せ配列を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検索するアルゴリズム。	Krogh, A. 他 (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. 他 (1988) Nucleic Acids Res. 26:320-322; Durbin, R. 他 (1998) Our World View, in a Nutshell, Cambridge Univ. Press. 1-350ページ。	PFAMヒット：確率値=1.0E-3 以下 シグナル ペプチド ヒット：スコア=0以上

【表９】

表7－2

プログラム	説明	参照文献	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.	標準化された質スコア≧特定のPrositeモチーフに対するGCG指定「HIGH」値 通常、スコア=1.4～2.1.
Phred	高度の感受性および確率のある自動シークエンサートレースを調べる塩基コーリングアルゴリズム。	Ewing, B. 他 (1998) Genome Res. 8:175-185; Ewing, B. および P. Green (1998) Genome Res. 8:186-194.	
Phrap	Smith-Waterman アルゴリズムの効率的実行に基づいたプログラムである、SWATとCrossMatchを含むPhils改訂構築プログラムで、配列相同性検索やDNA配列の構築に有用である。	Smith, T.F. および M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. および M.S. Waterman (1981) J. Mol. Biol. 147:195-197; Green, P., University of Washington, Seattle, WA.	スコア=120以上 一致した長さ=56以上
Consed	Phrap 構築の表示、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、分泌シグナルペプチドの存在を調べる重量マトリックス分析プログラム。	Nielson, H. 他 (1997) Protein Engineering 10:1-6; Claverie, J.M. および S. Audic (1997) CABIOS 12:431-439.	スコア=3.5以上
TMAP	重量マトリックスを使ってタンパク質配列上の膜貫通部分を描写し、方向を決定するプログラム。	Persson, B. および P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. および P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	隠れマルコフモデル(HMM) を使ってタンパク質配列上の膜貫通部分を描写し、方向を決定するプログラム。	Sonnhammer, E.L. 他 (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol., Biol., Glasgow 他, 編集, The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, 175-182ページ。 175-182.	
Motifs	Prositeで定義されたものと一致する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, バージョン9, M51-59ページ, Genetics Computer Group, Madison, WI	

【国際公開パンフレット】

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 August 2002 (01.08.2002)

PCT

(10) International Publication Number
WO 02/059323 A2

(51) International Patent Classification⁷: C12N 15/52, 9/00, 15/63, C07K 16/40, C12Q 1/68, G01N 33/50, 33/53, A61K 38/17

(21) International Application Number: PCT/US01/48575

(22) International Filing Date:
13 December 2001 (13.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/255,963 15 December 2000 (15.12.2000) 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)*: **LEE, Ernestine, A.** [US/US]; 624 Kains Street, Albany, CA 94706 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US).

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

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, 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 02/059323 A2

(54) **Title:** AMINOACYL TRNA SYNTHETASES

(57) **Abstract:** The invention provides human aminoacyl tRNA synthetases (ATRS) and polynucleotides which identify and encode ATRS. 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 ATRS.

AMINOACYL TRNA SYNTHETASES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of aminoacyl tRNA synthetases and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative and autoimmune/inflammatory disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of aminoacyl tRNA synthetases.

BACKGROUND OF THE INVENTION

Correct translation of the genetic code depends upon each amino acid forming a linkage with the appropriate transfer RNA (tRNA). The aminoacyl-tRNA synthetases (aaRSs) are essential proteins found in all living organisms. The aaRSs are responsible for the activation and correct attachment of an amino acid with its cognate tRNA, as the first step in protein biosynthesis. Prokaryotic organisms have at least twenty different types of aaRSs, one for each different amino acid, while eukaryotes usually have two aaRSs, a cytosolic form and a mitochondrial form, for each different amino acid. The 20 aaRS enzymes can be divided into two structural classes. Class I enzymes add amino acids to the 2' hydroxyl at the 3' end of tRNAs while Class II enzymes add amino acids to the 3' hydroxyl at the 3' end of tRNAs. Each class is characterized by a distinctive topology of the catalytic domain. Class I enzymes contain a catalytic domain based on the nucleotide-binding 'Rossman fold'. In particular, a consensus tetrapeptide motif is highly conserved (Prosite Document PDOC00161, Aminoacyl-transfer RNA synthetases class-I signature). Class I enzymes are specific for arginine, cysteine, glutamic acid, glutamine, isoleucine, leucine, methionine, tyrosine, tryptophan, and valine. Class II enzymes contain a central catalytic domain, which consists of a seven-stranded antiparallel ß-sheet domain, as well as N- and C- terminal regulatory domains. Class II enzymes are separated into two groups based on the heterodimeric or homodimeric structure of the enzyme; the latter group is further subdivided by the structure of the N- and C-terminal regulatory domains (Hartlein, M. and S. Cusack (1995) J. Mol. Evol. 40:519-530). Class II enzymes are specific for alanine, asparagine, aspartic acid, glycine, histidine, lysine, phenylalanine, proline, serine, and threonine.

Certain aaRSs also have editing functions. IleRS, for example, can misactivate valine to form Val-tRNA^{Ile}, but this product is cleared by a hydrolytic activity that destroys the mischarged product. This editing activity is located within a second catalytic site found in the connective polypeptide 1 region (CP1), a long insertion sequence within the Rossman fold domain of Class I enzymes (Schimmel, P. et al. (1998) FASEB J. 12:1599-1609). AaRSs also play a role in tRNA processing. It

has been shown that mature tRNAs are charged with their respective amino acids in the nucleus before export to the cytoplasm, and charging may serve as a quality control mechanism to insure the tRNAs are functional (Martinis, S.A. et al. (1999) EMBO J. 18:4591-4596).

In addition to their function in protein synthesis, specific aminoacyl tRNA synthetases also play roles in cellular fidelity, RNA splicing, RNA trafficking, apoptosis, and transcriptional and translational regulation. For example, human tyrosyl-tRNA synthetase can be proteolytically cleaved into two fragments with distinct cytokine activities. The carboxy-terminal domain exhibits monocyte and leukocyte chemotaxis activity as well as stimulating production of myeloperoxidase, tumor necrosis factor-α, and tissue factor. The N-terminal domain binds to the interleukin-8 type A receptor and functions as an interleukin-8-like cytokine. Human tyrosyl-tRNA synthetase is secreted from apoptotic tumor cells and may accelerate apoptosis (Wakasugi, K., and P. Schimmel (1999) Science 284:147-151). Mitochondrial Neurospora crassa TyrRS and S. cerevisiae LeuRS are essential factors for certain group I intron splicing activities, and human mitochondrial LeuRS can substitute for the yeast LeuRS in a yeast null strain. Certain bacterial aaRSs are involved in regulating their own transcription or translation (Martinis, supra). Several aaRSs are able to synthesize diadenosine oligophosphates, a class of signalling molecules with roles in cell proliferation, differentiation, and apoptosis (Kisselev, L.L et al. (1998) FEBS Lett. 427:157-163; Vartanian, A. et al. (1999) FEBS Lett. 456:175-180).

Autoantibodies against aminoacyl-tRNAs are generated by patients with autoimmune diseases such as rheumatic arthritis, dermatomyositis and polymyositis, and correlate strongly with complicating interstitial lung disease (ILD) (Freist, W. et al. (1999) Biol. Chem. 380:623-646; Freist, W. et al. (1996) Biol. Chem. Hoppe Seyler 377:343-356). These antibodies appear to be generated in response to viral infection, and coxsackie virus has been used to induce experimental viral myositis in animals.

Comparison of aaRS structures between humans and pathogens has been useful in the design of novel antibiotics (Schimmel, supra). Genetically engineered aaRSs have been utilized to allow site-specific incorporation of unnatural amino acids into proteins in vivo (Liu, D.R. et al. (1997) Proc. Natl. Acad. Sci. USA 94:10092-10097).

The discovery of new aminoacyl tRNA synthetases, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative and autoimmune/inflammatory disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of aminoacyl tRNA synthetases.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, aminoacyl tRNA synthetases, referred to collectively as "ATRS" and individually as "ATRS-1." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising the amino acid of SEQ ID NO:1, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO:1, c) a biologically active fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, and d) an immunogenic fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising the amino acid sequence of SEQ ID NO:1, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO:1, c) a biologically active fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, and d) an immunogenic fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1. In one alternative, the polynucleotide encodes a polypeptide of SEQ ID NO:1. In another alternative, the polynucleotide has the sequence of SEQ ID NO:2.

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 the amino acid sequence of SEQ ID NO:1, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO:1, c) a biologically active fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, and d) an immunogenic fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1. 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 the amino acid sequence of SEQ ID NO:1, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO:1, c) a biologically active fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, and d) an immunogenic fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1. 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 the amino acid

sequence of SEQ ID NO:1, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO:1, c) a biologically active fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, and d) an immunogenic fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising the polynucleotide sequence of SEQ ID NO:2, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to the polynucleotide sequence of SEQ ID NO:2, 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 the polynucleotide sequence of SEQ ID NO:2, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to the polynucleotide sequence of SEQ ID NO:2, 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 the polynucleotide sequence of SEQ ID NO:2, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to the polynucleotide sequence of SEQ ID NO:2, 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 the amino acid sequence of SEQ ID NO:1, b) a polypeptide comprising a naturally occurring amino acid sequence at

least 90% identical to the amino acid sequence of SEQ ID NO:1, c) a biologically active fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, and d) an immunogenic fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises the amino acid sequence of SEQ ID NO:1. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional ATRS, 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 the amino acid sequence of SEQ ID NO:1, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO:1, c) a biologically active fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, and d) an immunogenic fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1. 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 ATRS, 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 the amino acid sequence of SEQ ID NO:1, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO:1, c) a biologically active fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, and d) an immunogenic fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1. 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 ATRS, 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 the amino acid sequence of SEQ ID NO:1, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO:1, c) a biologically active fragment of

a polypeptide having the amino acid sequence of SEQ ID NO:1, and d) an immunogenic fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1. 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 the amino acid sequence of SEQ ID NO:1, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO:1, c) a biologically active fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, and d) an immunogenic fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1. 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 the polynucleotide sequence of SEQ ID NO:2, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising the polynucleotide sequence of SEQ ID NO:2, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to the polynucleotide sequence of SEQ ID NO:2, 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 the polynucleotide sequence of SEQ ID NO:2, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to the polynucleotide

sequence of SEQ ID NO:2, 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

"ATRS" refers to the amino acid sequences of substantially purified ATRS 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 ATRS. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of ATRS either by directly interacting with ATRS or by acting on components of the biological pathway in which ATRS participates.

An "allelic variant" is an alternative form of the gene encoding ATRS. 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 ATRS include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as ATRS or a polypeptide with at least one functional characteristic of ATRS. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding ATRS, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding ATRS. 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 ATRS. 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 ATRS 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 ATRS. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of ATRS either by directly interacting with ATRS or by acting on components of the biological pathway in which ATRS 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 ATRS 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 ATRS, 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 ATRS or fragments of ATRS 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 ATRS or the polynucleotide encoding ATRS 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:2 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:2, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:2 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:2 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:2 and the region of SEQ ID NO:2 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 is encoded by a fragment of SEQ ID NO:2. A fragment of SEQ ID NO:1 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1. For example, a fragment of SEQ ID NO:1 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1. The precise length of a fragment of SEQ ID NO:1 and the region of SEQ ID NO:1 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 ATRS 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 ATRS 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 ATRS. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of ATRS.

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 ATRS 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 ATRS.

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

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

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

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR 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 ATRS, nucleic acids encoding ATRS, 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 aminoacyl tRNA synthetases (ATRS),

the polynucleotides encoding ATRS, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative and autoimmune/inflammatory disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte 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 aminoacyl tRNA synthetases. For example, SEQ ID NO:1 is 50% identical to mouse valyl-tRNA synthetase (GenBank ID g4590328) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.5e-271, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a tRNA synthetase class I (I, L, M and V) domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ

ID NO:1 is a valyl-tRNA synthetase. The algorithms and parameters for the analysis of SEQ ID NO:1 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. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:2 or that distinguish between SEQ ID NO:2 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL*_XXXXXX_*N_1*_*N_2*_YYYYY_*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 *YYYYY* 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 polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX_gAAAAA_gBBBBB_*1*_N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, *gAAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, *gBBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq

sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, *gBBBBB*).

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

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

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

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

The invention also encompasses ATRS variants. A preferred ATRS 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 ATRS amino acid sequence, and which contains at least one functional or structural characteristic of ATRS.

The invention also encompasses polynucleotides which encode ATRS. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NO:2, which encodes ATRS. The polynucleotide sequences of SEQ ID NO:2, 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 ATRS. 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 ATRS. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising the sequence of SEQ ID NO:2 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the nucleic acid sequence of SEQ ID NO:2. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of ATRS.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding ATRS. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding ATRS, 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 ATRS 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 ATRS. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of ATRS.

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 ATRS, 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 ATRS, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode ATRS and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring ATRS under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding ATRS 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 ATRS 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 ATRS and ATRS 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 ATRS 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:2 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 ATRS 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 ATRS may be cloned in recombinant DNA molecules that direct expression of ATRS, 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 ATRS.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter ATRS-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 ATRS, 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 ATRS 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, ATRS 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 ATRS, 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 ATRS, the nucleotide sequences encoding ATRS or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding ATRS. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding ATRS. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding ATRS 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 ATRS 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 ATRS. 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 ATRS. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding ATRS 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 ATRS 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 ATRS are needed, e.g. for the production of antibodies, vectors which direct high level expression of ATRS 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 ATRS. 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 ATRS. Transcription of sequences encoding ATRS 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 ATRS 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 ATRS 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 ATRS in cell lines is preferred. For example, sequences encoding ATRS 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), ß glucuronidase and its substrate ß-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 ATRS is inserted within a marker gene sequence, transformed cells containing sequences encoding ATRS can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding ATRS 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 ATRS and that express ATRS 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 ATRS 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 ATRS 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 ATRS include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding ATRS, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic

agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding ATRS 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 ATRS may be designed to contain signal sequences which direct secretion of ATRS 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 ATRS 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 ATRS protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of ATRS 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 ATRS encoding sequence and the heterologous protein sequence, so that ATRS 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 ATRS 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.

ATRS of the present invention or fragments thereof may be used to screen for compounds that specifically bind to ATRS. At least one and up to a plurality of test compounds may be screened for specific binding to ATRS. 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 ATRS, 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 ATRS 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 ATRS, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing ATRS or cell membrane fractions which contain ATRS are then contacted with a test compound and binding, stimulation, or inhibition of activity of either ATRS 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 ATRS, either in solution or affixed to a solid support, and detecting the binding of ATRS 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.

ATRS of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of ATRS. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for ATRS activity, wherein ATRS is combined with at least one test compound, and the activity of ATRS in the presence of a test compound is compared with the activity of ATRS in the absence of the test compound. A change in the activity of ATRS in the presence of the test compound is indicative of a compound that modulates the activity of ATRS. Alternatively, a test compound is combined with an in vitro or cell-free system comprising ATRS under conditions suitable for ATRS activity, and the

assay is performed. In either of these assays, a test compound which modulates the activity of ATRS 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 ATRS 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 ATRS 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 ATRS 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 ATRS 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 ATRS, e.g., by secreting ATRS 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 ATRS and aminoacyl tRNA synthetases. In addition, examples of tissues

expressing ATRS can be found in Table 6. Therefore, ATRS appears to play a role in cell proliferative and autoimmune/inflammatory disorders. In the treatment of disorders associated with increased ATRS expression or activity, it is desirable to decrease the expression or activity of ATRS. In the treatment of disorders associated with decreased ATRS expression or activity, it is desirable to increase the expression or activity of ATRS.

Therefore, in one embodiment, ATRS 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 ATRS. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma.

In another embodiment, a vector capable of expressing ATRS 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 ATRS including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified ATRS 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 ATRS including, but not limited to, those provided above.

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

In a further embodiment, an antagonist of ATRS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of ATRS. Examples of such disorders include, but are not limited to, those cell proliferative and autoimmune/inflammatory disorders described above. In one aspect, an antibody which specifically binds ATRS 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 ATRS.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding ATRS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of ATRS 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 ATRS may be produced using methods which are generally known in the art. In particular, purified ATRS may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind ATRS. Antibodies to ATRS 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 ATRS 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 ATRS 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 ATRS 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 ATRS 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 ATRS-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 ATRS 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 ATRS and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering ATRS 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 ATRS. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of ATRS-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 ATRS epitopes, represents the average affinity, or avidity, of the antibodies for ATRS. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular ATRS epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the ATRS-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 ATRS, 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 ATRS-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 ATRS, 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 ATRS. 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 ATRS. (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 ATRS 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 ATRS expression or regulation causes disease, the expression of ATRS 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 ATRS are treated by constructing mammalian expression vectors encoding ATRS and introducing these vectors by mechanical means into ATRS-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 ATRS 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). ATRS 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 ATRS 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 ATRS expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding ATRS 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 ATRS to cells which have one or more genetic abnormalities with respect to the expression of ATRS. 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 ATRS to target cells which have one or more genetic abnormalities with respect to the expression of ATRS. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing ATRS 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 ATRS 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 ATRS into the alphavirus genome in place of the capsid-coding region results in the production of a large number of ATRS-coding RNAs and the synthesis of high levels of ATRS 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 ATRS 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 ATRS.

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 ATRS. 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 ATRS. 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 ATRS expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding ATRS may be therapeutically useful, and in the treatment of disorders associated with decreased ATRS expression or activity, a compound which specifically promotes expression of the polynucleotide encoding ATRS 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 ATRS 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 ATRS 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 ATRS. 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 ATRS, antibodies to ATRS, and mimetics, agonists, antagonists, or inhibitors of ATRS.

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 ATRS or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, ATRS 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 ATRS or fragments thereof, antibodies of ATRS, and agonists, antagonists or inhibitors of ATRS, 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 ATRS may be used for the diagnosis of disorders characterized by expression of ATRS, or in assays to monitor patients being treated with ATRS or agonists, antagonists, or inhibitors of ATRS. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for ATRS include methods which utilize the antibody and a label to detect ATRS 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 ATRS, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of ATRS expression. Normal or standard values for ATRS expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to ATRS under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of ATRS 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 ATRS 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 ATRS may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of ATRS, and to monitor regulation of ATRS levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding ATRS or closely related molecules may be used to identify nucleic acid sequences which encode ATRS. 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 ATRS, 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 ATRS 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:2 or from genomic sequences including promoters, enhancers, and introns of the ATRS gene.

Means for producing specific hybridization probes for DNAs encoding ATRS include the cloning of polynucleotide sequences encoding ATRS or ATRS 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 ATRS may be used for the diagnosis of disorders associated with expression of ATRS. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and an

autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma. The polynucleotide sequences encoding ATRS 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 ATRS expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding ATRS may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding ATRS 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 ATRS 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 ATRS, 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 ATRS, 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 ATRS 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 ATRS, or a fragment of a polynucleotide complementary to the polynucleotide encoding ATRS, 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 ATRS 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 ATRS 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 ATRS 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, ATRS, fragments of ATRS, or antibodies specific for ATRS 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 ATRS to quantify the levels of ATRS 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 ATRS 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 ATRS 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, ATRS, 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 ATRS 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 ATRS, or fragments thereof, and washed. Bound ATRS is then detected by methods well known in the art. Purified ATRS 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 ATRS specifically compete with a test compound for binding ATRS. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with ATRS.

In additional embodiments, the nucleotide sequences which encode ATRS 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. 60/255,963, 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). 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), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from 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:2. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative aminoacyl tRNA synthetases 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 amiinoacyl tRNA synthetases, the encoded polypeptides were analyzed by querying against PFAM models for aminoacyl tRNA synthetases. Potential aminoacyl tRNA synthetases were also identified by homology to Incyte cDNA sequences that had been annotated as aminoacyl tRNA synthetases. 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 ATRS Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:2 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:2 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 ATRS are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding ATRS. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of ATRS 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:2 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

desorbtion 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 ATRS-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring ATRS. 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 ATRS. 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 ATRS-encoding transcript.

XII. Expression of ATRS

Expression and purification of ATRS is achieved using bacterial or virus-based expression systems. For expression of ATRS 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 ATRS upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of ATRS 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 ATRS 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, ATRS 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 ATRS 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 ATRS obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, where applicable.

XIII. Functional Assays

ATRS function is assessed by expressing the sequences encoding ATRS 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 ATRS on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding ATRS 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 ATRS and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of ATRS Specific Antibodies

ATRS 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 ATRS 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-ATRS activity by, for example, binding the peptide or ATRS 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 ATRS Using Specific Antibodies

Naturally occurring or recombinant ATRS is substantially purified by immunoaffinity

chromatography using antibodies specific for ATRS. An immunoaffinity column is constructed by covalently coupling anti-ATRS 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 ATRS are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of ATRS (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/ATRS 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 ATRS is collected.

XVI. Identification of Molecules Which Interact with ATRS

ATRS, 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 ATRS, washed, and any wells with labeled ATRS complex are assayed. Data obtained using different concentrations of ATRS are used to calculate values for the number, affinity, and association of ATRS with the candidate molecules.

Alternatively, molecules interacting with ATRS 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).

ATRS 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 ATRS Activity

tRNA synthetase activity is measured as the aminoacylation of a substrate tRNA in the presence of [^{14}C]-labeled amino acid. ATRS is incubated with [^{14}C]-labeled amino acid and the appropriate cognate tRNA (for example, [^{14}C]valine and tRNA^{val}) in a buffered solution. ^{14}C-labeled product is separated from free [^{14}C]amino acid by chromatography, and the incorporated ^{14}C is quantified by scintillation counter. The amount of ^{14}C-labeled product detected is proportional to the activity of ATRS in this assay.

XVIII. Identification of ATRS Agonists and Antagonists

Agonists or antagonists of ATRS activation or inhibition may be tested using the assay described in section XVII. Agonists cause an increase in ATRS activity and antagonists cause a decrease in ATRS 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 Protein ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
60211391	1	60211391CD1	2	60211391CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
1	60211391CD1	g4590328	3.5e-271	Valyl-tRNA synthetase [Mus musculus] Snoek, M. and van Vugt, H. (1999) The sequence and organization of the mouse valyl-tRNA synthetase gene G7a/Bat6 located in the MHC class III region. Immunogenetics 49:468-470

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	60211391CD1	1063	S8 S40 S68 S72 S105 S288 S398 S447 S555 S613 S660 S674 S735 S741 S776 S832 S844 S904 S1019 S1039 T30 T336 T337 T567 T617 T1014 T1017 T1025 Y84 Y114 Y474	N147 N278	Transmembrane domains: S613-T634, V841-P858 (N terminus is cytosolic) tRNA synthetases class I (E, L, M and Y) Aminoacyl-transfer RNA synthetases class I signature PS00178: V148-A157, Q653-N665	TMAP HMMER-PFAM BLIMPS-BLOCKS PROFILESCAN
					Aminoacyl-transfer RNA synthetases class-I signature: P132-P181 Valyl-tRNA synthetase signature PR00986; P139-G150, H351-Y368, R467-W480, D579-E600, Y610-R628 SYNTHETASE AMINOACYLTRNA LIGASE BIOSYNTHESIS ATPBINDING LEUCYLTRNA VALYLTRNA LEURS LEUCINETRNA PD000389: W113-V393, W529-Y610, I416-W586 SYNTHETASE AMINOACYLTRNA VALYL TRNA VALYLTRNA VALINETRNA LIGASE VALRS BIOSYNTHESIS PD113286: Y104-E293 SYNTHETASE VALYLTRNA AMINOACYLTRNA VALINETRNA LIGASE VALRS BIOSYNTHESIS ATPBINDING 896AA PD000939: V391-I533 SYNTHETASE AMINOACYLTRNA LIGASE BIOSYNTHESIS ATPBINDING VALYLTRNA VALINETRNA VALRS ISOLEUCYLTRNA PD000476: P593-R926, P146-I166	PROFILESCAN BLIMPS-PRINTS BLAST-PRODOM BLAST-PRODOM BLAST-PRODOM BLAST-PRODOM
					AMINOACYL-TRANSFER RNA SYNTHETASES CLASS-I DM00514|P26640|506-1094: V312-P886 DM00514|P49696|455-1047: V312-A888 DM00514|P07806|352-933: V312-P886 DM00514|P28350|341-921: V312-S887 Aminoacyl-transfer RNA synthetases class-1 signature: P146-A157	BLAST-DOMO MOTIFS

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
2/ 6021139ICB1/ 3557	1-446, 1-678, 1-685, 1-755, 12-625, 22-649, 24-825, 29-812, 43-827, 46-297, 64-680, 65-829, 70-827, 79-770, 87-345, 89-827, 96-512, 96-552, 96-827, 97-343, 97-429, 106-299, 158-823, 192-463, 697-1299, 746-1286, 746-1367, 804-1117, 823-1245, 823-1300, 823-1348, 823-1358, 849-1422, 1045-1715, 1092-1647, 1109-1447, 1109-1561, 1109-1594, 1109-1680, 1109-1752, 1109-1808, 1109-1838, 1109-1846, 1109-1920, 1115-1562, 1156-1753, 1169-1765, 1178-1774, 1178-1979, 1199-1845, 1201-1963, 1215-1710, 1217-1883, 1234-1878, 1260-1835, 1269-1734, 1279-1635, 1282-1825, 1290-2088, 1323-1885, 1327-1689, 1327-1870, 1339-1932, 1398-2046, 1492-2402, 1557-1600, 1557-2363, 1574-2374, 1577-2287, 1578-2311, 1599-2346, 1629-2378, 1638-2396, 1650-2475, 1694-2254, 1711-2400, 1721-2371, 1728-2304, 1756-2465, 1761-2359, 1763-2404, 1771-2324, 1773-2383, 1793-2417, 1805-2468, 1824-2459, 1834-2304, 1857-2557, 1863-2396, 1866-2596, 1872-2683, 1882-2679, 1891-2380, 1897-2528, 1901-2682, 1908-2430, 1908-2580, 1911-2387, 1911-2518, 1924-2556, 1926-2466, 1932-2700, 1989-2732, 1993-2695, 2019-2679, 2061-2481, 2061-2482, 2061-2534, 2061-2587, 2065-2711, 2095-2703, 2096-2705, 2098-2792, 2107-2595, 2111-2733, 2122-2738, 2147-2703, 2152-2733, 2173-2726, 2181-2591, 2184-2801, 2185-2747, 2193-3051, 2201-2711, 2207-2726, 2226-2747, 2234-2685, 2236-2809, 2240-2702, 2263-2736, 2283-2843, 2287-2721, 2288-2876, 2291-2737, 2297-2862, 2300-2736, 2309-3003, 2334-3160, 2350-2736, 2351-2878, 2354-2721, 2355-2900, 2372-3029, 2390-2928, 2393-2899, 2410-2919, 2429-2916, 2430-3265, 2442-3278, 2455-2904, 2459-3172, 2469-2801, 2474-3120, 2478-3034, 2490-3203, 2491-3146, 2517-3328, 2523-2982, 2526-3111, 2537-3102, 2565-3099, 2573-3227, 2582-3185, 2591-3271, 2597-3223, 2629-3262, 2635-3256, 2655-3218, 2669-3229, 2671-3329, 2692-3555, 2754-3377, 2762-3249, 2763-3252, 2766-3319, 2782-3482, 2783-3451, 2793-3501, 2827-3556, 2878-3407, 2885-3556, 2888-3406, 2892-3451, 2900-3385, 2900-3392, 2900-3397, 2911-3397, 2933-3556, 2953-3296, 2953-3556, 2967-3557, 2989-3489, 3001-3451, 3100-3220

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
2	6021139ICB1	BRAITDR03

Table 6

Library	Vector	Library Description
BRAITDR03	PCDNA2.1	This random primed library was constructed using RNA isolated from allocortex; cingulate posterior 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.

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.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences:* Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, 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.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs:* fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, 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 score≥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.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

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

2. An isolated polypeptide of claim 1 comprising the amino acid sequence of SEQ ID NO:1.

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 the polynucleotide sequence of SEQ ID NO:2.

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 the amino acid sequence of

SEQ ID NO:1.

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 the polynucleotide sequence of SEQ ID NO:2,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to the polynucleotide sequence of SEQ ID NO:2,

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 the amino acid sequence of SEQ ID NO:1.

19. A method for treating a disease or condition associated with decreased expression of functional ATRS, 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 ATRS, 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 ATRS, 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 ATRS 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 ATRS 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 ATRS 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 the amino acid sequence of SEQ ID NO:1, 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 the amino acid sequence of SEQ ID NO:1.

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 the amino acid sequence of SEQ ID NO:1, 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 the amino acid sequence of SEQ ID NO:1.

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 the amino acid sequence of SEQ ID NO:1 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 the amino acid sequence of SEQ ID NO:1 in the sample.

45. A method of purifying a polypeptide comprising the amino acid sequence of SEQ ID NO:1 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 the amino acid sequence of SEQ ID NO:1.

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 polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:2.

<110> INCYTE GENOMICS, INC.
LEE, Ernestine A.
BAUGHN, Mariah R.

<120> AMINOACYL TRNA SYNTHETASES

<130> PI-0323 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/255,963
<151> 2000-12-15

<160> 2

<170> PERL Program

<210> 1
<211> 1063
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 60211391CD1

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

```
                260                 265                 270
Tyr Arg Asn His Gln Leu Val Asn Trp Ser Cys Ala Leu Arg Ser
                275                 280                 285
Ala Ile Ser Asp Ile Glu Val Glu Asn Arg Pro Leu Pro Gly His
                290                 295                 300
Thr Gln Leu Arg Leu Pro Gly Cys Pro Thr Pro Val Ser Phe Gly
                305                 310                 315
Leu Leu Phe Ser Val Ala Phe Pro Val Asp Gly Glu Pro Asp Ala
                320                 325                 330
Glu Val Val Val Gly Thr Thr Arg Pro Glu Thr Leu Pro Gly Asp
                335                 340                 345
Val Ala Val Ala Val His Pro Asp Asp Ser Arg Tyr Thr His Leu
                350                 355                 360
His Gly Arg Gln Leu Arg His Pro Leu Met Gly Gln Pro Leu Pro
                365                 370                 375
Leu Ile Thr Asp Tyr Ala Val Gln Pro His Val Gly Thr Gly Ala
                380                 385                 390
Val Lys Val Thr Pro Ala His Ser Pro Ala Asp Ala Glu Met Gly
                395                 400                 405
Ala Arg His Gly Leu Ser Pro Leu Asn Val Ile Ala Glu Asp Gly
                410                 415                 420
Thr Met Thr Ser Leu Cys Gly Asp Trp Leu Gln Gly Leu His Arg
                425                 430                 435
Phe Val Ala Arg Glu Lys Ile Met Ser Val Leu Ser Glu Arg Gly
                440                 445                 450
Leu Phe Arg Gly Leu Gln Asn His Pro Met Val Leu Pro Ile Cys
                455                 460                 465
Ser Arg Ser Gly Asp Val Ile Glu Tyr Leu Leu Lys Asn Gln Trp
                470                 475                 480
Phe Val Arg Cys Gln Glu Met Gly Ala Arg Ala Ala Lys Ala Val
                485                 490                 495
Glu Ser Gly Ala Leu Glu Leu Ser Pro Ser Phe His Gln Lys Asn
                500                 505                 510
Trp Gln His Trp Phe Ser His Ile Gly Asp Trp Cys Val Ser Arg
                515                 520                 525
Gln Leu Trp Trp Gly His Gln Ile Pro Ala Tyr Leu Val Val Glu
                530                 535                 540
Asp His Ala Gln Gly Glu Glu Asp Cys Trp Val Val Gly Arg Ser
                545                 550                 555
Glu Ala Glu Ala Arg Glu Val Ala Ala Glu Leu Thr Gly Arg Pro
                560                 565                 570
Gly Ala Glu Leu Thr Leu Glu Arg Asp Pro Asp Val Leu Asp Thr
                575                 580                 585
Trp Phe Ser Ser Ala Leu Phe Pro Phe Ser Ala Leu Gly Trp Pro
                590                 595                 600
Gln Glu Thr Pro Asp Leu Ala Arg Phe Tyr Pro Leu Ser Leu Leu
                605                 610                 615
Glu Thr Gly Ser Asp Leu Leu Leu Phe Trp Val Gly Arg Met Val
                620                 625                 630
Met Leu Gly Thr Gln Leu Thr Gly Gln Leu Pro Phe Ser Lys Val
                635                 640                 645
Leu Leu His Pro Met Val Arg Asp Arg Gln Gly Arg Lys Met Ser
                650                 655                 660
Lys Ser Leu Gly Asn Val Leu Asp Pro Arg Asp Ile Ile Ser Gly
                665                 670                 675
Val Glu Met Gln Leu Leu Gln Glu Lys Leu Arg Ser Gly Asn Leu
                680                 685                 690
Asp Pro Ala Glu Leu Ala Ile Val Ala Ala Ala Gln Lys Lys Asp
                695                 700                 705
Phe Pro His Gly Ile Pro Glu Cys Gly Thr Asp Ala Leu Arg Phe
                710                 715                 720
Thr Leu Cys Ser His Gly Val Gln Ala Gly Asp Leu His Leu Ser
                725                 730                 735
```

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

<210> 2
<211> 3557
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 60211391CB1

<400> 2

```
tggcgcggc ggctccaggg ccacgttcca gggtcgggct tggtggattc ctcagtcect 60
gccgccgcgg ggcgccctgg gatagcggcg gggcctcctg aacagatctc ggccccttte 120
caaacactcc tgatgcctca tttgcctctc gcctcttttc gaccaccatt ttgggggctg 180
aggcactcac ggggcctccc caggtttcac tccgtttcta cacagtcgga gcccatgga 240
tctcccatct cccggaggaa ccgtgaagcc aaacagaagc gcctgcgaga gaagcagggg 300
actctggagg ctgagatagc aggggagagc aagtcacctg cagaatccat taaggcctgg 360
aggcctaagg agttagtatt gtatgaaatc cctacgaaac ccggtgaaaa gaaagatgtc 420
tctgggcccc tgcctcctgc atacagcccc cgatatgttg aggctgcctg gtacccgtgg 480
```

tgggtacgag agggcttctt caaaccagaa tatcaggccc ggctgcccca agctacaggg 540
gagacctttt ccatgtgtat cccacctccc aatgtcactg gctccctgca cattggccac 600
gcactcacgg tggccataca ggatgccctc gtgcgctggc accggatgcg tggggatcaa 660
gtgctgtggg tccctggttc agatcatgca ggaattgcta cacaagctgt ggtggagaaa 720
caactgtgga aggaacgggg agtgaggaga catgagctga gccgggaggc cttccttagg 780
gaggtgtggc agtggaagga ggcgaaaggt ggagagatct gtgagcagct gcgagctctg 840
ggtgcctccc tggactggga tcgagagtgt tttaccatgg atgttggctc ctcagtggct 900
gtgactgaag cttttgtgcg gctctacaag gcgcggttgc tgtaccggaa ccatcagctt 960
gtcaactggt catgtgcttt aagatcagcc atctcggaca ttgaggtgga gaaccggccc 1020
ctgcctggcc acacacagct tcgactgcct ggctgcccca ccccgtgtc ttttggcctc 1080
ctattttctg ttgccttccc cgtggatgga gagcctgatg cagaggttgt ggtaggaacc 1140
acaaggccag agacgctgcc tggagatgtg gctgtggccg ttcatccaga cgactcgcga 1200
tacacacatc tacacgggcg acagcttcgt caccccttga tggggcagcc tcttcccctc 1260
atcacagact atgctgttca gccacatgtg ggcacggggg cagtgaaggt gactccagct 1320
cacagtcctg ccgatgctga gatgggggcc cgacatggct tgagcccctt gaatgtcatt 1380
gcggaggatg ggaccatgac ctccctctgc ggggactggt tgcagggtct tcaccggttt 1440
gtggcccggg aaaagataat gtctgtgctg agtgaacggg gcctattccg gggcctccag 1500
aaccacccca tggtactgcc catctgcagc cgttctgggg atgtgataga atacctgctg 1560
aagaaccagt ggtttgtccg ctgccaggaa atgggggccc gagctgccaa ggctgtggag 1620
tcgggggccc tggagctcag tccctccttc caccagaaga actggcagca ctggttttcc 1680
catattgggg actggtgtgt ctcccggcag ctgtggtggg gccatcagat tccagcctac 1740
ctggttgtag aggaccatgc gcagggagaa gaggactgtt gggtggttgg gcggtcagag 1800
gctgaggcca gagaggtagc agcggaactg acagggaggc caggggcaga gctgaccctg 1860
gagagggatc ctgatgtcct agacacatgg ttttcttctg ccctgttccc cttttctgcc 1920
ctgggctggc cccaagagac cccagacctt gctcgtttct acccctctgtc actttttggaa 1980
acgggcagcg accttctgct gttctgggtg ggccgcatgg tcatgttggg gacccagctc 2040
acagggcagc tgcccttcag caaggtgctt cttcatccca tggttcggga caggcagggc 2100
cggaagatga gcaagtccct ggggaatgtg ctggacccaa gagacatcat cagtgggggtg 2160
gagatgcagt tgctgcagga aaagctgaga agcggaaatt tggaccctgc agagctggcc 2220
attgtggctg cagcacagaa aaaggacttt cctcacggga tccctgagtg tgggacagat 2280
gccctgagat tcacactctg ctcccatgga gttcaggcgg gcgacttgca cctgtcagtc 2340
tctgaggtcc agagctgccg acatttctgc aacaagatct ggaatgctct tcgctttatc 2400
ctcaatgctt taggggagaa atttgtgcca cagcctgctg aggagctgtc tccctcctcc 2460
ccgatggatg cctggatcct gagccgcctt gccctggctg cccaggagtg tgagcggggc 2520
ttcctcaccc gagagctctc gctcgtcact catgccctgc accacttctg gcttcacaac 2580
ctctgtgacg tctacctgga ggctgtgaag cccgtgctgt ggcactcgcc ccgccccctg 2640
gggccccctc aggtcctgtt ctcctgcgct gacctcggcc tccgcctcct ggccccactg 2700
atgcccttcc tggctgaaga gctctggcag aggctgcccc ccaggcctgg ttgccccctt 2760
gccccagcca tctcggttgc cccctaccct agcgcctgca gcttggagca ctggcgccag 2820
ccagagctgg agcgggcgtt ctcccgggtc caagaggtcg tgcaggtgct aagggctctc 2880
cgagccacgt accagctcac caaagcccgg ccccgagtgc tgctgcagag ctcagagcct 2940
ggggaccagg gcctcttcga ggcttcttg gagccctgg gcaccctggg ctactgtggg 3000
gctgtgggcc tgttacccc aggcacagca gctcctccg gctgggccca ggctccactc 3060
agtgacacgg ctcaagtcta catggagctg cagggcctgg tggacccgca gatccagcta 3120
cctctgttag ccgcccgaag gtacaagttg cagaagcagc ttgatagcct cacagccagg 3180
acccatcag aaggggaggc agggactcag aggcaacaaa agctttcttc cctccagctg 3240
gaattgtcaa aactggacaa ggcagcctct cacctccggc agctgatgga tgagcctcca 3300
gcccaggga gcccggagct ctaactcatc atcccatca gttttcctcc ctctcagacc 3360
tgtctttgag gacaaacaga tttgtcagct gtcagggtgc agtgggacgt cagagactat 3420
gtggtccatc gccttcattg tgtaaatgag gacacagact ggcttggtcg cagtgactgt 3480
ggtgtccttg agatgctcac attactgccc ggcctgcctc ccacctggaa gtctgggaat 3540
gaggagattg agataaa 3557

【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
1 August 2002 (01.08.2002)

PCT

(10) International Publication Number
WO 02/059323 A3

(51) International Patent Classification⁷: C12N 15/52, 9/00, 15/63, C07K 16/40, C12Q 1/68, G01N 33/50, 33/53, A61K 38/17

(21) International Application Number: PCT/US01/48575

(22) International Filing Date:
13 December 2001 (13.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/255,963 15 December 2000 (15.12.2000) 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)*: **LEE, Ernestine, A.** [US/US]; 624 Kains Street, Albany, CA 94706 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US).

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

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, 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:
24 July 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 02/059323 A3

(54) Title: AMINOACYL TRNA SYNTHETASES

(57) Abstract: The invention provides human aminoacyl tRNA synthetases (ATRS) and polynucleotides which identify and encode ATRS. 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 ATRS.

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/48575

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/52 C12N9/00 C12N15/63 C07K16/40 C12Q1/68
G01N33/50 G01N33/53 A61K38/17

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K C12Q

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)
SEQUENCE SEARCH, EMBL, EPO-Internal, BIOSIS, CHEM ABS Data, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 01 90330 A (INCYTE GENOMICS INC ;PATTERSON CHANDRA (US); TRIBOULEY CATHERINE M) 29 November 2001 (2001-11-29) Seq. ID no 3: 99.83% identity with seq ID no 1 of application in 598 aa; Seq. ID no 7 : 99.7% identity with seq. ID no 2 of application in 2043 bp claims 1-52; tables 2,3 --- -/--	1-57

[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
27 January 2003	20/02/2003

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	Trommsdorff, M

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

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/48575

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	VILALTA ADRIAN ET AL: "Cloning, sequencing and expression of a cDNA encoding mammalian valyl-tRNA synthetase." GENE (AMSTERDAM), vol. 123, no. 2, 1993, pages 181-186, XP001121728 ISSN: 0378-1119 figures 1,3 & DATABASE EMBL 'Online! 13 November 1995 (1995-11-13) VILALATA ET AL.: "Human tansfer valyl-tRNA synthetase mRNA 3'end of cds" retrieved from EMBL Database accession no. M98326 abstract	1-57
Y	HSIEH S-L ET AL: "EVIDENCE THAT GENE G7A IN THE HUMAN MAJOR HISTOCOMPATIBILITY COMPLEX ENCODES VALYL-TRANSFER RNA SYNTHETASE" BIOCHEMICAL JOURNAL, vol. 278, no. 3, 1991, pages 809-816, XP009002689 ISSN: 0264-6021 figures 4,5	1-57
Y	DATABASE EMBL 'Online! 22 February 2000 (2000-02-22) SUGANO ET AL.: "Homo sapiens cDNA, highly similar to SYV_FUGRU Valyl-tRNA synthetase" Database accession no. AK000511 XP002224910 abstract	1-57
Y	SCHIMMEL PAUL ET AL: "Aminoacyl tRNA synthetases as targets for new anti-infectives" FASEB JOURNAL (FEDERATION OF AMERICAN SOCIETIES FOR EXPERIMENTAL BIOLOGY), BETHESDA, US, vol. 12, no. 15, December 1998 (1998-12), pages 1599-1609, XP002191843 ISSN: 0892-6638 cited in the application the whole document	1-57
Y	EP 0 909 818 A (SMITHKLINE BEECHAM CORP) 21 April 1999 (1999-04-21) page 18, line 20 -page 23, line 5; claims 1-20	1-57

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

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 01/48575

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 0190330	A	29-11-2001	AU 6340401 A WO 0190330 A2		03-12-2001 29-11-2001
EP 0909818	A	21-04-1999	EP 0909818 A1		21-04-1999

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

フロントページの続き

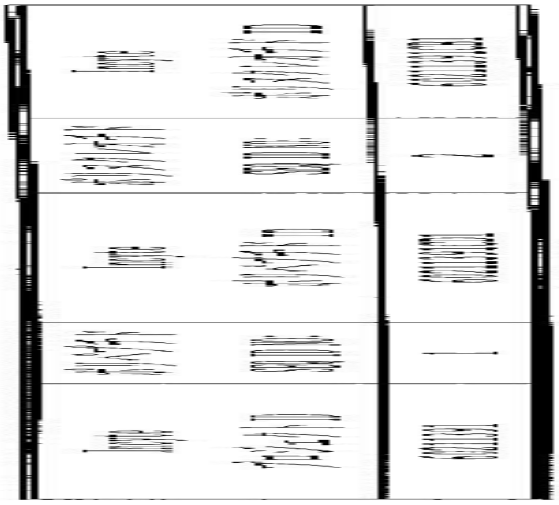
(51)Int.Cl.[7]	FI			テーマコード（参考）
Ａ６１Ｐ 35/00	Ａ６１Ｐ 37/02			４Ｂ０６５
Ａ６１Ｐ 37/02	Ａ６１Ｐ 43/00	１１１		４Ｃ０８４
Ａ６１Ｐ 43/00	Ｃ０７Ｋ 16/40			４Ｃ０８５
Ｃ０７Ｋ 16/40	Ｃ１２Ｍ 1/00		Ａ	４Ｈ０４５
Ｃ１２Ｍ 1/00	Ｃ１２Ｎ 1/15			
Ｃ１２Ｎ 1/15	Ｃ１２Ｎ 1/19			
Ｃ１２Ｎ 1/19	Ｃ１２Ｎ 1/21			
Ｃ１２Ｎ 1/21	Ｃ１２Ｎ 9/00			
Ｃ１２Ｎ 5/10	Ｃ１２Ｑ 1/68		Ａ	
Ｃ１２Ｎ 9/00	Ｇ０１Ｎ 33/15		Ｚ	
Ｃ１２Ｑ 1/68	Ｇ０１Ｎ 33/50		Ｚ	
Ｇ０１Ｎ 33/15	Ｇ０１Ｎ 33/53		Ｄ	
Ｇ０１Ｎ 33/50	Ｇ０１Ｎ 33/53		Ｍ	
Ｇ０１Ｎ 33/53	Ｇ０１Ｎ 33/566			
Ｇ０１Ｎ 33/566	Ｇ０１Ｎ 37/00	１０２		
Ｇ０１Ｎ 37/00	Ｃ１２Ｎ 5/00		Ａ	
	Ａ６１Ｋ 37/48			

(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 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

Ｆターム(参考) 2G045 AA34 AA35 BB01 BB20 BB29 BB51 CB01 DA13 DA14 DA36
FB02 FB03 FB05 FB08 FB12 GC15 JA00
4B024 AA01 AA11 BA07 BA44 CA01 FA02 GA11 HA12 HA20
4B029 AA23 BB20 CC03 FA01 FA09
4B050 CC03 DD11 EE10 LL01 LL03
4B063 QA01 QA18 QA20 QQ44 QQ79 QQ96 QR32 QR48 QR55 QS32
QS33 QS34 QX01
4B065 AA93Y AB01 AC14 AC20 BA02 CA27 CA44 CA46
4C084 AA02 AA06 AA07 AA17 BA01 BA08 BA22 BA23 DC01 NA14
ZB071 ZB072 ZB111 ZB112 ZB261 ZB262 ZC022
4C085 HH01 KA03 KA04 KA05 KA26 LL18
4H045 AA10 AA11 BA10 CA40 DA76 DA89 EA22 EA28 EA50 FA74

专利名称(译)	氨酰基tRNA合成酶		
公开(公告)号	JP2004537270A	公开(公告)日	2004-12-16
申请号	JP2002559808	申请日	2001-12-13
[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	リーアーンスティーンエイ ボーグンマライアアール		
发明人	リー、アーンスティーン·エイ ボーグン、マライア·アール		
IPC分类号	G01N33/50 A61K38/00 A61K38/43 A61K45/00 A61K49/00 A61P29/00 A61P35/00 A61P37/02 A61P43/00 C07K16/40 C12M1/00 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N9/00 C12N15/09 C12Q1/68 G01N33/15 G01N33/53 G01N33/566 G01N37/00		
CPC分类号	A61K38/00 A61P29/00 A61P35/00 A61P37/02 A61P43/00 C12N9/93		
FI分类号	C12N15/00.ZNA.A A61K45/00 A61K49/00.A A61P29/00 A61P35/00 A61P37/02 A61P43/00.111 C07K16/40 C12M1/00.A C12N1/15 C12N1/19 C12N1/21 C12N9/00 C12Q1/68.A G01N33/15.Z G01N33/50.Z G01N33/53.D G01N33/53.M G01N33/566 G01N37/00.102 C12N5/00.A A61K37/48		
F-TERM分类号	2G045/AA34 2G045/AA35 2G045/BB01 2G045/BB20 2G045/BB29 2G045/BB51 2G045/CB01 2G045/DA13 2G045/DA14 2G045/DA36 2G045/FB02 2G045/FB03 2G045/FB05 2G045/FB08 2G045/FB12 2G045/GC15 2G045/JA00 4B024/AA01 4B024/AA11 4B024/BA07 4B024/BA44 4B024/CA01 4B024/FA02 4B024/GA11 4B024/HA12 4B024/HA20 4B029/AA23 4B029/BB20 4B029/CC03 4B029/FA01 4B029/FA09 4B050/CC03 4B050/DD11 4B050/EE10 4B050/LL01 4B050/LL03 4B063/QA01 4B063/QA18 4B063/QA20 4B063/QQ44 4B063/QQ79 4B063/QQ96 4B063/QR32 4B063/QR48 4B063/QR55 4B063/QS32 4B063/QS33 4B063/QS34 4B063/QX01 4B065/AA93Y 4B065/AB01 4B065/AC14 4B065/AC20 4B065/BA02 4B065/CA27 4B065/CA44 4B065/CA46 4C084/AA02 4C084/AA06 4C084/AA07 4C084/AA17 4C084/BA01 4C084/BA08 4C084/BA22 4C084/BA23 4C084/DC01 4C084/NA14 4C084/ZB071 4C084/ZB072 4C084/ZB111 4C084/ZB112 4C084/ZB261 4C084/ZB262 4C084/ZC022 4C085/HH01 4C085/KA03 4C085/KA04 4C085/KA05 4C085/KA26 4C085/LL18 4H045/AA10 4H045/AA11 4H045/BA10 4H045/CA40 4H045/DA76 4H045/DA89 4H045/EA22 4H045/EA28 4H045/EA50 4H045/FA74		
优先权	60/255963 2000-12-15 US		
外部链接	Espacenet		

摘要(译)

本发明提供了人氨酰基tRNA合成酶（ATRS）和鉴定和编码ATRS的多核苷酸。 本发明还提供表达载体，宿主细胞，抗体，激动剂和拮抗剂。 本发明还提供了用于诊断，治疗或预防与ATRS异常表达有关的疾病的方法。