

(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2004-526444

(P2004-526444A)

(43) 公表日 平成16年9月2日(2004.9.2)

(51) Int.Cl. ⁷	F I	テーマコード (参考)
C 1 2 N 15/09	C 1 2 N 15/00	2 G O 4 5
A 6 1 K 31/7088	A 6 1 K 31/7088	4 B O 2 4
A 6 1 K 38/00	A 6 1 K 39/395	4 B O 6 3
A 6 1 K 39/395	A 6 1 K 39/395	4 B O 6 4
A 6 1 K 48/00	A 6 1 K 48/00	4 B O 6 5
審査請求 未請求 予備審査請求 未請求 (全 100 頁) 最終頁に続く		

(21) 出願番号	特願2002-571510 (P2002-571510)	(71) 出願人	301005050
(86) (22) 出願日	平成14年3月7日 (2002.3.7)		インサイト・ゲノミックス・インコーポレイテッド
(85) 翻訳文提出日	平成15年9月8日 (2003.9.8)		アメリカ合衆国カリフォルニア州94304・パロアルト・ポータードライブ 3160
(86) 国際出願番号	PCT/US2002/007053		
(87) 国際公開番号	W02002/072596	(74) 代理人	100089266
(87) 国際公開日	平成14年9月19日 (2002.9.19)		弁理士 大島 陽一
(31) 優先権主張番号	09/802, 520	(72) 発明者	ラル、ブリーティ・ジー
(32) 優先日	平成13年3月9日 (2001.3.9)		アメリカ合衆国カリフォルニア州95056・サンタクララ・ビーオーボックス 5142
(33) 優先権主張国	米国 (US)		
		最終頁に続く	

(54) 【発明の名称】 STEAP関連タンパク質

(57) 【要約】

本発明はSTEAP関連タンパク質をコードするcDNAを提供する。本発明はまた、cDNAの利用のために、自身の断片、補体、及び変異体を提供し、コードされたタンパク質の利用のために、とりわけ前立腺肥大及び前立腺癌のような前立腺細胞増殖性疾患の診断及び治療のために自身の一部及び抗体を提供する。本発明はまた、タンパク質及び遺伝子組換モデルシステムの作成のための発現ベクターおよび宿主細胞をも提供する。

【特許請求の範囲】

【請求項 1】

SEQ ID NO:1のアミノ酸配列、若しくはその補体を有するタンパク質をコードする核酸配列を含む単離されたcDNA。

【請求項 2】

核酸配列を含む単離されたcDNAであって、

(a) SEQ ID NO:2、若しくはその補体と、

(b) SEQ ID NO:3-9から選択されたSEQ ID NO:2の断片、及びその補体と、

(c) SEQ ID NO:10から選択されたSEQ ID NO:2の変異体、もしくはその補体とから選択される単離されたcDNA。

10

【請求項 3】

前記請求項 1 及び標識化部分のcDNA若しくはその補体を有する組成物。

【請求項 4】

前記請求項 1 のcDNAを含むベクター。

【請求項 5】

前記請求項 4 のベクターを有する宿主細胞。

【請求項 6】

タンパク質を生成するためにcDNAを用いる方法であって、

(a) タンパク質発現の条件下で、前記請求項 5 の宿主細胞を培養する過程と、

(b) 前記宿主細胞の培養から前記タンパク質を回収する過程とを含む方法。

20

【請求項 7】

サンプル中の核酸の発現を検出するのにcDNAを用いる方法であって、

(a) 前記サンプルの核酸に前記請求項 3 の組成物をハイブリダイズし、ハイブリダイゼーション複合体を生成する過程と、

(b) ハイブリダイゼーション複合体形成物と標準物質とを比較する過程であって、前記比較過程が前記サンプル中の前記cDNAの発現を示すことを特徴とする過程とを含む方法。

【請求項 8】

さらに、ハイブリダイゼーションの前に、前記サンプルの前記核酸を増幅させる過程を含むことを特徴とする請求項 7 に記載の方法。

【請求項 9】

前記組成物が基質に付着されることを特徴とする請求項 7 の方法。

30

【請求項 10】

前記cDNAが、前記標準物質と比較される場合に差動的に発現され、前立腺肥大もしくは前立腺癌の指標となることを特徴とする請求項 7 の方法。

【請求項 11】

複数の分子若しくは混合物をスクリーニングするためにcDNAを用いる方法であって、

特異的な結合を可能とするための条件下で、前記請求項 1 のcDNAと複数の分子若しくは混合物とを組み合わせる過程と、

特異的な結合を検出し、それにより前記cDNAに特異的に結合する分子若しくは混合物を同定する過程とを含む方法。

40

【請求項 12】

前記分子若しくは混合物が、DNA分子、RNA分子、ペプチド核酸、人工クロモソーム構造、ペプチド、転写要素、リプレッサー、及び調節分子から選択されることを特徴とする請求項 11 に記載の方法。

【請求項 13】

前記請求項 6 の方法で作成された精製されたタンパク質若しくはその一部であって、

(a) SEQ ID NO:1のアミノ酸配列と、

(b) SEQ ID NO:1の抗原性エピトープと、

(c) SEQ ID NO:1の生物学的活性部位とから選択される精製されたタンパク質若しくはその一部。

50

【請求項 14】

前記請求項 13 のタンパク質及び薬剤キャリア (pharmaceutical carrier) を含む組成物。

【請求項 15】

少なくとも一つのリガンドを同定するべく、複数の分子若しくは混合物をスクリーニングするためにタンパク質を用いる方法であって、
特異的な結合を可能とするための条件下で、前記請求項 13 のタンパク質と前記分子若しくは混合物とを結合させる過程と、
特異的な結合を検出し、それにより前記タンパク質に特異的に結合するリガンドを同定する過程とを含む方法。

10

【請求項 16】

前記分子若しくは混合物が、DNA分子、RNA分子、ペプチド核酸、ペプチド、タンパク質、模倣 (mimetic)、アゴニスト、アンタゴニスト、抗体、免疫グロブリン、阻害剤、及び薬剤から選択されることを特徴とする請求項 15 に記載の方法。

【請求項 17】

抗体を調整し精製するべくタンパク質を用いる方法であって、
(a) 抗体反応を誘発する条件下で、前記請求項 15 のタンパク質で動物を免疫する過程と、
(b) 動物の抗体を単離する過程と、
(c) 前記タンパク質を基質に付着させる過程と、
(d) 前記タンパク質に対する特異的な結合が可能となるような条件下で、前記基質と単離した抗体とを接触させる過程と、
(e) 前記抗体を前記タンパク質より解離させ、それによって精製された抗体を獲得する過程とを含む方法。

20

【請求項 18】

前記請求項 17 の方法によって生成された抗体。

【請求項 19】

タンパク質の発現に関する疾患もしくはコンディションを診断するための、抗体を用いた方法であって、
(a) 前記請求項 18 の抗体とサンプルとを結合させ、前記抗体：タンパク質複合体を形成する過程と、
(b) 複合体フォームーションと標準物質とを比較する課程であって、ここで前記比較が前記サンプル中の前記タンパク質の発現を示すことを特徴とする課程とを有する方法。

30

【請求項 20】

発現が前立腺肥大もしくは前立腺癌の診断となることを特徴とする前記請求項 19 に記載の方法。

【請求項 21】

前期請求項 18 の抗体の前記特異性を有する単クローン性抗体を調整するための方法であって、

(a) 抗体反応を誘発する条件下で SEQ ID NO:1 のタンパク質で動物を免疫する過程と、
(b) 前記動物より抗体生成細胞を単離する過程と、
(c) 培養内で前記抗体生成細胞と不死化細胞とを融合させ、ハイブリドーマ細胞を生成する単クローン性抗体を形成する過程と、
(d) 前記ハイブリドーマ細胞を培養する過程と、
(e) 培養より単クローン性抗体を単離する過程とを含む方法。

40

【請求項 22】

前記請求項 21 の方法によって生成された単クローン性抗体。

【請求項 23】

抗体を用いて、タンパク質を免疫精製するための方法であって、

(a) 基質に請求項 18 の抗体を付着させる過程と、

50

- (b) 抗体：タンパク質複合体を形成可能とする条件下で、タンパク質を有するサンプルに対して前記抗体を露出させる過程と、
(c) 前記複合体より前記タンパク質を分離させる過程と、
(d) 前記精製タンパク質を回収する過程とを含む方法。

【請求項 24】

請求項 18 の抗体及び標識部分を含む組成物。

【請求項 25】

請求項 18 の抗体及び薬剤を含む組成物。

【発明の詳細な説明】

【技術分野】

10

【0001】

本発明は、STEAP関連タンパク質をコードするcDNAに関し、また、とりわけ前立腺過形成もしくは前立腺癌のような前立腺細胞増殖性疾患に於ける診断及び治療でcDNA及びコードされたタンパク質を用いることに関する。

【背景技術】

【0002】

生物間系統発生的関係は何回も実証されてきた。そして広範な原核細胞および真核細胞生物の研究により分子、生化学的メカニズム、生理学的メカニズム、代謝経路の多少の漸進的な進化が示されてきた。進化ストレスが異なるが、線虫、ハエ、ラット、ヒトのタンパク質には共通の化学的及び構造的特徴があり、一般的に同じ細胞機能を行う。構造と機能が公知である生物種での核酸とタンパク質配列の比較は、ヒト配列の研究を加速する。またヒトの症状、疾患または障害に対する診断と薬剤をテストするモデル系の発達を可能にする。

20

【0003】

前立腺癌は、50歳より年齢が上の男性に共通の悪性のものであり、その発生率は高年齢であるほど高い。米国では、新規に前立腺癌と診断されたおよそ132,000人の患者がおり、毎年33,000以上が死亡している。ひとたび前立腺内部に癌細胞が発生すると、より迅速な増殖へ、テストステロンによって刺激される。精巣の除去は、癌の迅速な増殖及び転位の双方を間接的に減少させる。95%を越える前立腺癌が、前立腺腺房で起こる腺癌である。残りの5%は、扁平細胞及び移行上皮癌の間で分割され、双方が前立腺管若しくは前立腺のその他の部位で起こる。

30

【0004】

ほとんどの癌と同様に、究極的には多段階進行を経て前立腺癌が発生し、活動的、転位性のフェノタイプの結果として、結果的には活動的で転位性のフェノタイプとなる。腫瘍の進行における初期ステップは、正常な管腔及び/又は基本的上皮性細胞の過剰増殖を伴い、それらは初期段階の腫瘍へと過形成し、また進化することとなる。初期段階の腫瘍は、前立腺内に局在化されるが、最終的には、とりわけ骨、脳、肺へと転移する。それら腫瘍の約80%は、前立腺上皮細胞の成長を制御するアンドロゲン治療に応じて残存する。しかし、その進行状態において、癌の成長はアンドロゲン非依存性となり、この状態のためには現在治療法が知られていない。

40

【0005】

前立腺癌のための初期診断マーカーは、前立腺特異型抗原(PSA)である。PSAは、もっぱら前立腺上皮性細胞により生成される組織特異型セリンプロテアーゼである。PSAの量は、前立腺上皮性細胞の数及び量に関連しており、結果的にPSAのレベルは異常な前立腺成長の優れたインジケータである。前立腺癌を患う男性は、診断に先立っての指数関数的増加によって追従されたPSAレベルに於いて、初期の直線的増加を示す。しかし、PSAレベルはまた、炎症、アンドロゲン、及びその他成長要素のような要素に影響を受けるので、複数の科学者及び臨床家は、PSAレベルにおける変化が前立腺癌の個々のケースを検出するのに有用であると主張する。

【0006】

50

現在の癌研究の分野は、前立腺癌のための潜在的な治療ターゲットはもちろんのこと、マーカーの追加的予測も提供する。複数の成長因子が、腫瘍の発生、成長、及び進行における重大な役割を果たすことが示されてきている。成長因子である、上皮性成長因子(EGF)、線維芽細胞成長因子(FGF)、及びトランスフォーミング成長因子(TGF)は、とりわけ腫瘍の成長及び進行の初期段階における超増殖前立腺上皮性細胞はもちろんのこと、通常の成長に於いて重要であり、また、様々な方法でそれら細胞におけるシグナル経路に影響を与える(Linら(1999) Cancer Res 59:2891-2897; Putzら(1999) Cancer Res 59:227-233)。成長因子のTGF-ファミリーは、通常、ヒトの癌に於いて高いレベルで発現し、多くの場合の高い発現レベルは、悪性かつ乏しい残存率の、進行型ステージに関連する(Gold(1999) Crit Rev Oncog 10:303-360)。最終的には、アンドロゲン独立型はもちろんのこと前立腺癌(LNCap)のアンドロゲン依存型の双方、前立腺癌の進行に関連する遺伝子発現パターンの研究に於いて有用である疾患のホルモン不応期(PC3及びDU-145)、及びそれら発現遺伝子の細胞治療の効果を示すヒト細胞系が存在する(Chung(1999) Prostate 38:199-207)。

10

20

30

40

50

【0007】

前立腺の6つの膜貫通上皮性抗原(STEAP)は、前立腺特異性の細胞表面マーカーである(Hubertら(1999) Proc Natl Acad Sci 96:14523-14528)。STEAPは、全長が339のアミノ酸であり、6つの予想膜貫通領域を有する。通常及び癌性の前立腺組織において、また複数の前立腺癌誘導細胞株において、それは高く発現する。発現のレベルは、アンドロゲンの存在に対して鈍感である。免疫染色は、STEAPが、分泌上皮の細胞間の連絡に集中する前立腺細胞の原形質膜に位置していることを示している。STEAPのような細胞表面アンチゲンは、抗体治療、癌のワクチン、及び前立腺癌の治療のための画像診断に於いて有用である。

【0008】

STEAP関連タンパク質をコードするcDNAの発見は、とりわけ前立腺過形成もしくは前立腺癌のような前立腺細胞増殖性疾患に於ける診断及び治療で有用な成分を提供することで本技術分野の要求に応えるものである。

【発明の開示】

【発明の効果】

【0009】

本発明は、とりわけ前立腺過形成もしくは前立腺癌のような前立腺細胞増殖性疾患の診断・治療に有用なSTEAP関連タンパク質(STEAPRP)をコードするcDNAの発見に基づいている。

【0010】

本発明は、SEQ ID NO: 1のアミノ酸配列を有するタンパク質をコードする核酸配列を有する単離されたcDNAを提供する。SEQ ID NO:2の核酸配列を有する本発明はまた単離されたcDNAもしくはその補体、SEQ ID NOs:3-9から選択されたSEQ ID NO:2の断片、及びSEQ ID NO:2及び10の変異体も提供する。加えて本発明は、STEAPRPをコードするcDNAもしくはその補体を有する組成物、基質、及びプローブを提供する。更に本発明は、このようなcDNAを含むベクター、このようなベクターを含む宿主細胞、及びこのようなcDNAを用いてSTEAPRP作製する方法を提供する。本発明はさらにSTEAPRPをコードするcDNAを有するベクターを備えた遺伝子組み換え株化細胞及び生物を提供する。また本発明はさらにSEQ ID NO:2-10を含む群より選択された断片、変異体、もしくは補体を提供する。或る実施態様では、本発明は少なくとも一つのそれらの断片、変異体、もしくはその補体を含む基質を提供する。第2の実施態様では、本発明は、検出法、スクリーニング法、及び精製法に用いることが可能なcDNAもしくはその補体を含むプローブを提供する。更なる実施態様では、プローブは一本鎖相補RNAまたはDNA分子である。

【0011】

本発明はまた、cDNAを用いて、サンプルの核酸の発現変動を検出する方法であって、前記プローブと前記核酸とをハイブリダイズさせてハイブリダイゼーション複合体を形成し、

そのハイブリダイゼーション複合体を標準と比較し、その比較によりサンプルにおける前記cDNAの発現変動が分かる方法を提供する。或る実施態様では、この検出方法は、ハイブリダイゼーションの前に前記核酸を増幅するステップを含む。別の実施態様では、この検出方法はcDNAの異なった発現が、とりわけ前立腺過形成もしくは前立腺癌のような前立腺細胞増殖性疾患の診断に用いられることを示している。また別の側面では、cDNA、断片、変異体、若しくは補体はエレメント若しくはアレイを含んでも良い。

【0012】

本発明はまた、cDNA、その断片、または変異体、またはその相補体を用いて、ライブラリ即ち複数の分子または化合物をスクリーニングしてそのcDNAと特異的に結合する少なくとも1つのリガンドを同定する方法であって、特異的な結合が許容される条件下で、そのcDNAを前記分子または化合物と結合させるステップと、前記cDNAに対する特異的な結合を検出して、前記cDNAと特異的に結合するリガンドを特定するステップとを含む方法を提供する。一実施態様では、このような分子または化合物として、アダプター、DNA分子、RNA分子、ペプチド核酸、人工染色体作製物、ペプチド、転写因子、リプレッサー、及び調節分子が選択される。

10

【0013】

本発明はまた、SEQ ID NO:1のアミノ酸配列と、SEQ ID NO:1のアミノ酸配列と55%の同一性を有する変異体と、SEQ ID NO:1の抗原エピートープと、SEQ ID NO:1の生物学的に活性部分とからなる群から選択される精製されたタンパク質またはその一部を提供する。本発明はまた、医薬用担体に関連する前記タンパク質を含む組成物を提供する。本発明は更に、とりわけ前立腺過形成もしくは前立腺癌のような前立腺細胞増殖性疾患を有する被験者を治療するためにSTEAPRPを用い、そのような治療を必要とする患者に生成したタンパク質を含む合成物を投与する方法を提供する。本発明は更に、タンパク質を用いてライブラリ即ち複数の分子または化合物をスクリーニングして少なくとも1つのリガンドを特定する方法であって、特異的な結合が許容される条件下で、そのタンパク質を前記分子または化合物と結合させるステップと、特異的な結合を検出して、前記タンパク質と特異的に結合するリガンドを特定するステップとを含む方法を提供する。ある実施態様においては、分子または化合物はDNA分子、RNA分子、ペプチド核酸、ペプチド、タンパク質、擬態、アゴニスト、アンタゴニスト、抗体、免疫グロブリン、インヒビター、及び薬剤から選択される。別の実施態様では、とりわけ前立腺過形成もしくは前立腺癌のような前立腺細胞増殖性疾患の患者を治療するために前記リガンドを使用する。

20

30

【0014】

本発明は、タンパク質を用いて、被検サンプルをそのタンパク質に特異的に結合する抗体に対してスクリーニングする方法であって、前記被検サンプルから抗体を単離するステップと、特異的な結合が許容される条件下で、この単離した抗体を前記タンパク質とを結合させるステップと、結合した前記タンパク質から抗体を解離するステップと、抗体の存在或いは存在量からとりわけ乳ガン、卵巣癌、及び腎臓癌のような癌、またはとりわけ前立腺過形成もしくは前立腺癌のような前立腺細胞増殖性疾患の患者を診断できる抗体の量を既知の標準値と比較するステップとを含むスクリーニング法を提供する。

【0015】

本発明はまた、タンパク質を用いて抗体を作製及び精製する方法であって、抗体反応が起こる条件下でこのタンパク質によって動物を免疫するステップと、動物抗体を単離するステップと、基質にこのタンパク質を付着させるステップと、このタンパク質への特異的な結合が許容される条件下で、この基質を単離した抗体と接触させること、このタンパク質から抗体を分離して精製した抗体を得ることを含む方法を提供する。

40

【0016】

本発明は、とりわけ前立腺過形成もしくは前立腺癌のような前立腺細胞増殖性疾患において発現するタンパク質に特異的に結合する精製された抗体を提供する。本発明はまた、とりわけ前立腺過形成もしくは前立腺癌のような前立腺細胞増殖性疾患診断する方法であって、抗体を混合して結合抗体量を既知の指標と比較するステップを含み、それによって、

50

とりわけ前立腺過形成もしくは前立腺癌のような前立腺細胞増殖性疾患の存在を証明する診断方法を提供する。本発明は更に、抗体を用いてとりわけ前立腺過形成もしくは前立腺癌のような前立腺細胞増殖性疾患を治療する方法であって、精製された抗体を含む医薬組成物をそのような治療が必要な患者に投与することを含む治療方法を提供する。

【0017】

本発明はまた、哺乳動物のゲノムDNAの中に異種性マーカー遺伝子を挿入して、内在性ポリヌクレオチドの発現を阻害する方法を提供する。本発明はまた、cDNAを用いて哺乳動物モデル系を作製する方法であって、SEQ ID NO:2-10から選択されるcDNAを含むベクターを作製するステップと、そのベクターで胚性幹細胞を形質転換するステップと、形質転換された胚性幹細胞を選択するステップと、この形質転換された胚性幹細胞を哺乳動物胚盤胞の中に微量注入して、キメラ胚盤胞を形成するステップと、偽妊娠メスにキメラ胚盤胞を導入し、このメスが、生殖細胞系の中にcDNAを含むキメラ子孫を出産し、そのキメラ哺乳動物を交配して同型接合哺乳動物モデル系を作製するステップとを含む哺乳動物モデル系作製方法を提供する。

10

【発明を実施するための最良の形態】

【0018】

(本発明の記載について)

本発明は、ここに開示した特定の装置及び材料、方法に限定されず、その実施形態を変更できることを理解されたい。また、ここで用いられる用語は、特定の実施例のみを説明する目的で用いるものであり、後述の請求の範囲によってのみ限定され、本発明の範囲を限定することを意図したものではないということも理解されたい。本明細書及び請求の範囲において単数形を表す「或る」、「その(この等)」は、文脈で明確に示していない場合は複数形を含むことに注意されたい。従って、例えば「或る宿主細胞」は当業者には周知の複数の宿主細胞を含む。

20

【0019】

本明細書で用いた全ての科学技術用語は、別の方法で定義されていない限り、本発明の属する技術分野の一般的な技術者が普通に解釈する意味と同じである。本明細書に記載の全ての文献は、本発明に関連して使用する可能性のある文献に記載された細胞系、プロトコル、試薬、ベクターを記述し開示するために引用した。従来の特許を引用したからと言って、本発明の新規性が損なわれると解釈されるものではない。

30

【0020】

(定義)

「STEAPRP」は、天然、合成、半合成或いは組換え体など任意の哺乳動物種(ウシ、イヌ、マウス、ヒツジ、ブタ、げっ歯類、サルそして好ましくはヒトを含む)から得られる精製されたタンパク質を指す。

【0021】

「アレイ」は、基質上の少なくとも2つのcDNAの規則正しい配列を指す。cDNAの少なくとも1つが調節または標準を表す。そして他方が目的の診断もしくは治療のcDNA体を表す。基質上の2から約4万のcDNAの構成により、各々のcDNAと少なくとも一つのサンプル核酸の間で形成される各標識ハイブリダイゼーション複合体の大きさ及びシグナル強度を確実に個別に区別できる。

40

【0022】

配列表の核酸分子の「相補配列」は、完全長配列に完全に相補的なcDNAを指す。また高いストリンジェンシー条件下でcDNA或いはmRNAにハイブリダイズする。

【0023】

「cDNA」とは、単離したポリヌクレオチド、核酸分子或いはその任意の断片または相補配列を指す。それは組換えまたは合成された二本鎖または一本鎖であり、コード配列及び/または非コード5'及び3'配列を有し、一般的にはイントロンを欠く。

【0024】

「タンパク質をコードするcDNA」という語は、当分野で周知の分析により同定された保存

50

された領域、モチーフ或いはドメインをコードする配列と密接にアラインメントする核酸配列を指す。これらの分析には、保存された領域内における同一性を特定するBLAST (Basic Local Alignment Search Tool)が含まれる(Altschul (1993) J Mol Evol 36: 290-300、Altschul ら(1990) J Mol Biol 215:403-410)。

【0025】

「成分」という語は、ポリヌクレオチドを含み、また標識化部分もしくは薬学的キャリアに関連の精製されたタンパク質を含む。

【0026】

「誘導体」とは化学修飾されたcDNA或いはタンパク質を指す。cDNAの誘導体化にはクエオシン (queosine) 或いはヒポキサンチンなどの類似体等非従来型塩基の置換が含まれ得る。これらの置換は当分野で周知である。タンパク質の誘導体化にはアセチル基、アシル基、アルキル基、アミノ基、ホルミル基またはモルホリン基による水素の置換が含まれる。分子誘導体は天然分子の生物学的活性を保持するが、長い寿命或いは強化された活性などの長所を付与し得る。

10

【0027】

「発現変動」とは、サンプル中の転写されたメッセンジャーRNA或いは翻訳されたタンパク質の量の存在の有無、もしくは少なくとも二倍の変更により検出される増加即ちアップレギュレーション又は存在、或いは減少即ちダウンレギュレーション又は不在を指す。

【0028】

「障害」とは、cDNA及びSTEAPRPが異なって発現されている場合の症状、疾患、または症候群を指す。そのような疾患には、とりわけ前立腺過形成もしくは前立腺癌のような前立腺細胞増殖性疾患が含まれる。

20

【0029】

「断片」とは長さが約50から約4000塩基の連続したヌクレオチド鎖を指す。断片は、関連する核酸分子を同定するためにPCRまたはハイブリダイゼーション技術に、或いはリガンドをスクリーニングするための結合アッセイで使用され得る。そのようなリガンドは、複製、転写、または翻訳を調節する治療として有用である。

【0030】

「ハイブリタイゼーション複合体」とは、例えば5'-A-G-T-C-3'と3'-T-C-A-G-5'との塩基対などのように1つの分子のプリンが相補的な分子ピリミジンと水素結合して、cDNAとサンプルの核酸との間で形成される。ハイブリタイゼーションのコンディション、相補性の度合、およびヌクレオチド類似体の使用が、ハイブリダイゼーション反応の効率とストリエンジェンシーに影響を与える。

30

【0031】

「標識化成分」とは、cDNAまたはタンパク質に取り付けられるもしくは組み込まれ得るよりむしろ、可視標識もしくは放射性標識を指す。これに限定するものではないが、可視標識には、アントシアニン、緑色蛍光タンパク質(GFP)、グルクロニダーゼ、ルシフェラーゼ、Cy3及びCy5、又はそれに類似するものを含む。放射性標識は、放射性形態のヨウ素、亜リン酸、イオウ、及びそれに類似するものを含む。

【0032】

「リガンド」とは、ポリヌクレオチド或いはタンパク質のエピトープに特異結合するあらゆる物質、分子、または化合物を指す。そのようなリガンドは、ポリヌクレオチドまたタンパク質の活性を安定化或いは調節し、核酸、タンパク質、炭水化物、脂肪、及び脂質を含む無機および/または有機物質から構成され得る。

40

【0033】

「オリゴヌクレオチド」とは、長さが約18から約60ヌクレオチドの一本鎖分子を指す。そしてハイブリダイゼーションや増幅技術において、または複製、転写、翻訳の調節において使用され得る。概ね等しい用語としては、アンプライマー、プライマー、及びオリゴマーが挙げられる。

【0034】

50

「部分」とは、あらゆる目的に使用されるタンパク質の任意の部分を指すが、特にリガンドのスクリーニングまたは抗体の生産に用いられるエピトープを指す。

【0035】

タンパク質の「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、及び蛋白分解性切断等が含まれ得る。これらのプロセスは、合成或いは生化学的に生じ得る。生化学的修飾は、細胞の位置、細胞型、pH、及び酵素環境などによって異なる。

【0036】

「プローブ」とは、サンプル中の少なくとも1つの核酸にハイブリダイズするcDNAを指す。標的が一本鎖である場合、プローブは相補的一本鎖である。プローブは、サザン法、ノーザン法、insitu法、ドットプロット法、及びアレイなどを含むハイブリダイゼーション反応またはスクリーニングアッセイで使用するためにレポーター分子で標識化することができる。

10

【0037】

「タンパク質」とは、ポリペプチド或いはその任意の部分を指す。タンパク質の「部分」とは、少なくとも1つの生物学的活性を保持し得るアミノ酸配列の長さ、PFAMまたはPRINTS分析によって同定されるドメイン、或いはPROTEAN プログラム(DNASTAR, Madison WI)のKyte-Doolittle アルゴリズムを使用して同定されたタンパク質の抗原エピトープを指す。「オリゴペプチド」は、抗体を産生させるために融合タンパク質の一部として使用される約5残基から約15残基までのアミノ酸配列である。

20

【0038】

「精製された」とは、自然環境から分離され、自然環境で会合していた他の化合物が約60%から約90%まで分離したあらゆる分子や化合物を指す。

【0039】

「サンプル」は、核酸、タンパク質、及び抗体等を含むとして、その最も広い意味で用いられる。サンプルは体液、細胞調製の可溶性分画、細胞が成長する培地のアリコット、染色体、細胞小器官、或いは細胞から単離または抽出された膜、溶液中のまたは基質に固定されたゲノムDNA、RNA、またはcDNAと、細胞、組織、組織プリント、フィンガープリント、口内細胞、皮膚、または髪等を含み得る。

【0040】

「特異的な結合」とは、構造、特に分子側鎖に依存する2つの分子間での特殊な正確な相互作用を指す。例えば、調節タンパク質のDNA分子の主溝への挿入またはタンパク質のエピトープとアゴニスト、アンタゴニストまたは抗体との間の結合がある。

30

【0041】

配列に適用される「類似性」とは、Smith-Waterman アルゴリズム(Smith及びWaterman (1981) J Mol Biol 147:195-197)或いはBLAST2 (Altschul ら (1997) Nucleic Acids Res 25:3389-3402)などの標準化されたアルゴリズムを用いてアラインメントされた少なくとも2つの配列間で一致する分子或いは残基の定量化(通常は%)を指す。BLAST2は、アラインメントを最適化するために配列の1つでギャップを挿入するまた2つの配列をより有意に比較できる標準化された再現性のある方法で使用され得る。とりわけタンパク質では、類似性は保存的置換における同一性よりも大きく、例えばロイシンもしくはイソロイシンのためのバリンは報告されたパーセンテージを計算することでカウントされる。保存的であると考えられる置換は、当業者には周知である。

40

【0042】

「基質」とは、cDNAまたはタンパク質が結合する任意の固体または半固体の支持体を指すものであって、膜、フィルタ、チップ、スライド、ウエハ、ファイバー、磁性または非磁性ビーズ、ゲル、毛管、またはその他のチューブ、プレート、ポリマー、微小粒子が含まれ、穴、溝、ピン、チャンネル、孔等、様々な表面形態を有する。

【0043】

「変異体」とは、cDNAまたはそのcDNAがコードするタンパク質の認識できる変異した分子

50

を指す。スプライス変異体は、スコアが少なくとも100、最も好ましいのは少なくとも400であるBLASTスコアにより決定し得る。対立遺伝子変異体は、cDNAに対して高い同一性のパーセントを有し、100塩基に付き約3塩基が異なり得る。「一塩基多型」(SNP)とは、欠失、挿入、または置換による単一塩基による変異を指す。この変異は、保存的(プリンからプリン)或いは非保存的(プリンからピリミジン)であり得、コードされたアミノ酸に変異が起こる可能性がある。

【0044】

(発明)

本発明は、STEAPRPをコードするcDNAの発見と、とりわけ前立腺過形成もしくは前立腺癌のような前立腺細胞増殖性疾患の性質決定、診断、治療における、cDNAまたはその断片、およびタンパク質またはその部分をそのまま使用するか組成物として使用することに基づくものである。

10

【0045】

本発明のSTEAPRPをコードする核酸は、アミノ酸配列のためのコンピュータサーチを用い、脳歯状核cDNAライブラリ(BRAWTDR02)よりインサイトクローン710809内で最初に同定された。SEQ ID NO:2(7492448CB 1)は、追従するオーバーラッピング及び/又は拡張された核酸配列(SEQ ID NOs:3-9):インサイトクローン、7100809HI (BRAWTDR02)、6912820J1 (PITUDIR01)、4647117F6 (PROSTUT20)、7004364H1 (COLNFEC01)、70351677DI (SG0000177)、4108079HI (PROSBPT07)、及び4669848HI1(SDITNOT24)に由来した。表1は、組織カテゴリーにかかる転写物の発現を示すものであり、転写物の最も高い存在量が男性の生殖組織内に見られる(42%)。STEAPRPは、このカテゴリー内の前立腺組織において排他的に発現する。表2は、とりわけ腺繊維筋腫過形成前立腺上皮異常増殖、及び腺癌の患者より摘出した、前立腺組織に於ける転写物の発現を示している。STEAPRPは、腺繊維筋腫過形成の患者よりの前立腺組織ライブラリ(PROSNOT19、PROSDIT01、PROSNOT20、及びPROSNOT06)、上皮内異常増殖の患者よりの前立腺組織ライブラリ(PROETMP06)、及び腺癌の患者よりの前立腺組織ライブラリ(PROSTUT18、PROSTUS20、PROSTUT04、PROSTUT21、PROSTUS19、及びPROSTUT12)内で発現した。表3は、PrEC非腫瘍形成性前立腺上皮性細胞に対するヒトLNCaP前立腺癌におけるSTEAPRPの発現変動を示すものである。細胞は、実験中で異なった条件下で成長した。飢餓細胞(Starved cells)は、成長ファクター及びホルモンが不在の場合は、基本培地内で成長した。リッチ媒体が、成長ファクター及び栄養分を含み成長を促進させた。STEAPRPは、すべての成長状況下で、PrECに対するLNCaP癌細胞内に増加した発現を示す。転写物は、それ故、とりわけ前立腺過形成や前立腺癌のような前立腺細胞増殖疾患のための診断アッセイに有用である。およそヌクレオチド1から50までのcDNA断片もまた、診断アッセイにおいて有用である。

20

30

【0046】

ある実施例では、本発明は図1A乃至1Gに示すようにSEQ ID NO: 1のアミノ酸配列を含むポリヌクレオチドを包含する。STEAPRPは長さが490のアミノ酸である。また、STEAPRPは、N256に存在する一つの潜在的N-グリコシル化部位と、T32に存在する一つの潜在的環状AMPもしくはGMP依存タンパク質キナーゼリン酸化部位と、S12、T77、S100、S128、S197、及びS348に存在する6つのカゼインキナーゼIIリン酸化部位と、S9、T46、S197、S328、及びS455に存在する5つの潜在的プロテインキナーゼCリン酸化部位と、Y423に存在する一つの潜在的チロシンキナーゼリン酸化部位を有する。PFAM解析は、T32よりL136までのSTEAPRP領域がKTN NAD結合ドメインに類似することを示している。KTN NAD結合ドメインは、酢酸カリウム、ホスホフルクトキナーゼ、及び複数のトランスポーターを含む複数のタンパク質中に見られる。BLOCKS解析は、G34からK64までのSTEAPRPの領域が細菌型フィトエンデヒドロゲナーゼに類似しており、T32からV56までの領域はピリジンヌクレオチドジスルフィドクラス2オキシドレダクターゼに類似しており、T32からF70は6ホスホグルコン酸デヒドロゲナーゼに類似していることを示している。PRINTS解析は、V317からY331までのSTEAPRP領域がフタル酸ジオキシゲナーゼレダクターゼファミリーサインに類似しており、V33からI47の領域がアドレノドキシンレダクターゼファミリーサインに類似していること

40

50

を示している。これらモチーフの存在は、酸化還元反応に於けるSTEAPRPのために可能な機能を示す。STEAPRPの隠れマルコフモデルは、T210からP238、E253からQ281、C301からS328、M359からI379、F391からL411、及びF426からI454までの6つの膜貫通ドメイン存在を示し、M359からN387のシグナルペプチド領域の存在を示す。図2A乃至2Cに示すように、STEAPRPはヒトSTEAP (g6572948; SEQ ID NO: 11)と化学的及び物理的同一性を有している。とりわけ、STEAPRPとSTEAPとは、43%の同一性及び6つの膜貫通領域を共有している。有用な抗原エピドープは、およそG59 からD75、D234からK249、及びS455からT478に広がっており、STEAPRPの生化学的に活性な部位はおよそT32からL136に広がっている。STEAPRPに特異的に結合する抗体は、診断アッセイに有用であり、とりわけ前立腺過形成もしくは前立腺癌のような前立腺細胞増殖性疾患を認識する。

10

【0047】

TRPをコードするcDNAの哺乳動物変異体は、デフォルトのパラメータ及びZOOSEQデータベース(Incyte Genomics)と共にBLAST2を用いて同定された。これら好適な変異体は、以下の表に示すように約85%の同一性を有する。第1列は、ヒトcDNAのためのSEQ ID(SEQ IDH)を示しており、第2列は、変異体cDNAのためのSEQ ID(SEQ IDvar)を示しており、第3列は変異体cDNAのためのクローン番号(Clonevar)を示しており、第4列は、ライブラリの名前を示しており、第5列は、ヒトcDNAに対する変異体の配列を示しており、そして第6列はヒトcDNAに対する同一性のパーセンテージを示している。

【0048】

SEQ ID _H	SEQ ID _{var}	クローン _{var}	ライブラリ名	N _H 配列	同一性
2	10	702819778T1	RATSNON03	285-607	85%

20

【0049】

SEQ ID NOs:10のこれらcDNAは、遺伝子組み換え細胞株若しくは生物体を生成するのにとりわけ有用である。

【0050】

遺伝的コード悪化の結果として、STEAPRPをコードする複数のcDNA、既知のcDNAに対最小の類似性を有する複数のベアリング(bearing)、及び自然発生遺伝子が生成されても良いことは当業者が認識している。そのようなわけで、本発明は、可能なコドンに基づいたコンディションを選択することによってなされうるcDNAのすべての可能なバリエーションを意図するものである。これらバリエーションは、自然発生したSTEAPRPをコードするポリヌクレオチドに適用されたような通常型の三重遺伝子コードに関連してなされ、それらすべてのバリエーションは特異的に開示されていると考えられる。

30

【0051】

SEQ ID NOs:2-10のcDNAは、SEQ ID NO:2およびサンプル内の関連分子間で同定及び識別可能する合成、増幅、及びスクリーニング技術において用いられても良い。ほ乳類のcDNAは、遺伝子組み換え細胞株若しくは生物を生成するのに用いられ得る。それらは、とりわけ前立腺過形成もしくは前立腺癌のような前立腺細胞増殖性疾患のモデルシステムであり、潜在的な治療の毒性及び効果がテストされても良い。毒物学の学習、臨床試験、及び被験者/患者の処置のプロフィールは、cDNA、タンパク質、抗体、分子および化合物を用いて実行及びモニターされても良い。前記分子及び化合物は、本発明の分子及びcDNAやタンパク質を用いて識別されたものである。

40

【0052】

(本発明の特徴及び使用)

cDNAライブラリ

ここに開示する特定の実施例では、当分野で周知の方法を用いて哺乳動物の細胞及び組織からmRNAを単離し、これを用いてcDNAライブラリを作製する。上記したインサイト社クローンは、以降に記載する実施例に示された哺乳動物cDNAライブラリから単離された。本発

50

明の代表的な3つのライブラリの作製方法をす。コンセンサス配列は、Phrap (P. Green, University of Washington, Seattle WA)及びAUTOASSEMBLERアプリケーション(Applied Biosystems, Foster City CA)などのコンピュータプログラムを用いて、インサイト社クローンを含む断片、伸長、及び/またはショットガン配列から化学的かつ/または電子的に構築した。5'及び3'配列の確認の後、少なくとも一つのTRPをコードする代表cDNAが試薬とされる。

【0053】

シークエンシング

核酸をシークエンシングする方法は当分野で周知であり、そのような方法を用いて本発明の任意の実施例を実施することができる。これらの方法は、DNAポリメラーゼIであるクレノウフラグメント、SEQUENASE、Taq DNAポリメラーゼおよび熱耐性T7 DNAポリメラーゼ(Amersham Pharmacia Biotech (APB), Picataway NJ)、或いはELONGASE増幅システム(Life Technologies, Rockville MD)に用いられるような校正エクソヌクレアーゼとポリメラーゼとの組み合わせを用いることができる。配列の準備は、MICROLAB 2200(Hamilton, Reno NV)、及びDNA ENGINEサーマルサイクラー(PTC200; MJ Research, Watertown MA)などの装置を用いて自動的に行うのが望ましい。シークエンシングに用いる装置には、ABI 3700、377または373DNAシークエンシングシステム(Applied Biosystems)、及びMEGABACE 1000 DNAシークエンシングシステム(APB)等がある。当分野で周知の様々なアルゴリズムを用いて、シークエンシングした配列を解析することができる。これらのアルゴリズムは、Ausubelら(1997;Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7)及びMeyers(1995;Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853)に記載されている。

10

20

30

【0054】

ショットガン・シークエンシングは、目的の特定クローンインサートの配列を完成させるべく用いられ得る。ショットガン理論は、様々なサイズのセグメントに対するオリジナルインサートの破壊と、それら断片のベクターへのクローニングとをランダムに伴う。断片は、オリジナルインサートの全体配列が既知となるまで重複末端を用いて配列及び再構築される。ショットガンシークエンシング方法は当分野で周知であり、熱耐性DNAポリメラーゼや非熱耐性DNAポリメラーゼ、及び目的のcDNAに隣接する代表的な領域から選択されたプライマーを用いる。当分野で周知のCONSED(Gordon (1998) Genome Res. 8:195-202)などの様々なアルゴリズムやプログラムを用いて、組立てが未終了の配列(組立てが不完全な配列)を調べる。ベクターやキメラ配列、または欠失配列を含む汚染配列を除去して、組立てが未終了の配列を完全な配列に組み立てる。

【0055】

核酸配列の伸長

本発明の配列は、当分野で周知の様々なPCR法を用いた方法で伸長することができる。例えば、XL-PCRキット(Applied Biosystems)及び入れ子プライマー(nested primer)、市販のcDNAまたはゲノムDNAライブラリ用いてヌクレオチド配列を伸長することが可能である。全てのPCR系の方法に用いることができるように、プライマーは、OLIGO 4.06プライマー分析ソフトウェア(Molecular Biology Insights, Cascade CO)等の市販のソフトウェアを用いて、ヌクレオチドの長さが約22~30個、GC含量が約50%以上、約55~68の温度で標的配列とアニールするように設計することが可能である。調節エレメントを復活させるために配列を伸長する場合は、cDNAライブラリよりゲノムライブラリを用いる方が良い。

40

【0056】

ハイブリダイゼーション

cDNA及びその断片は、様々な目的のための様々なハイブリダイゼーション技術に用いることができる。プローブは、5'調節領域や非保存領域(すなわち、タンパク質の保存された触媒ドメインをコードする5'若しくは3'のヌクレオチド)等のユニークな領域から作製可能であり、STEAPRP、対立遺伝子の変種、若しくは関連分子をコードする天然の分

50

子を同定するためのプロトコルに用いることができる。通常は一本鎖であるDNA若しくはRNAからなるこのプローブは、任意のこの核酸配列、配列番号2 - 10と少なくとも50%の配列同一性を有することが望ましい。ハイブリダイゼーションプローブは、レポータ分子の存在下でのPCR増幅、オリゴ標識化、ニックトランスレーション法、または末端標識化を利用して作製することができる。このcDNAまたはその断片を含むベクターを用いて、RNAポリメラーゼ及び標識したヌクレオチドを加えて *in vitro* で mRNA プローブを作製することができる。これらの方法はAPBが販売するキットを用いて行うことができる。

【0057】

ハイブリダイゼーションのストリンジェンシー（厳密性）は、プローブのGC含量、塩濃度、及び温度によって決まる。特に、塩濃度を下げる、またはハイブリダイゼーションの温度を上げて、ストリンジェンシーを高めることができる。ハイブリダイゼーションは、低いストリンジェンシーの緩衝液（5×SSC、1%のドデシル硫酸ナトリウム（SDS））で、60℃で行うことができるが、核酸配列間に不適正塩基対を含む複合体の形成を許容し得る。続く洗浄は、45℃（中程度のストリンジェンシー）或いは68℃（高いストリンジェンシー）の何れかの温度、0.2×SSC、0.1% SDSなどの高いストリンジェンシーで行う。高いストリンジェンシーでは、ハイブリダイゼーション複合体は、完全に相補的な核酸分子部分のみが安定して保持される。ある膜系のハイブリダイゼーションにおいて、好ましくは35%、最も好ましくは50%のホルムアミドをハイブリダイゼーション溶液に加えて、ハイブリダイゼーションを行う温度を下げたり、または、SarkosylやTriton X-100 (Sigma-Aldrich, St. Louis Mo.)などの界面活性剤及び変性したサケ精子DNAなどのブロッキング試薬を用いてバックグラウンドシグナルを低減することが可能である。ハイブリダイゼーションの条件や要素の選択については当分野で周知であり、Ausubel(前出)及びSambrook他(1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY.に記載されている。

【0058】

当分野で周知の方法でマイクロアレイを準備して分析することができる。オリゴヌクレオチドもしくはcDNAは、ハイブリダイゼーションプローブや標的として用いられてもよく、同時に極めて多数の遺伝子の発現レベルをモニタリングし、遺伝子変異体、突然変異及びSNP（一塩基多型）を同定する。アレイが、遺伝子機能の解明や、症状及び疾患、または障害における遺伝子原理の解明や、症状及び疾患、障害の診断または治療、治療薬の開発、並びにこれらの治療薬の活性のモニタリングに用いられても良い。（例えば、Brennan他(1995) USPN 5,474,796; Schena他(1996) Proc. Natl. Acad. Sci. 93:10614-10619; Heller他(1997) Proc. Natl. Acad. Sci. 94:2150-2155; 及び Heller他(1997) USPN 5,605,662を参照）。

【0059】

ハイブリダイゼーションプローブはまた、天然のゲノム配列のマッピングに有用である。このプローブは、特定の染色体、染色体の特定の領域、人工染色体作製物にハイブリダイズすること可能である。そのような人工染色体作製物には、ヒト人工染色体（HAC）や酵母人工染色体（YAC）、細菌人工染色体（BAC）、細菌P1作製物、若しくは単一クロモソームcDNAライブラリを含む。

【0060】

発現

STEAPRPをコードする複数のcDNAの何れか一つをベクターにクローニングして、このタンパク質若しくはその一部を宿主細胞で発現させることができる。この核酸配列を、DNAシャフリング（Stemmer and Cramer (1996) USPN 5,830,721）や部位特異的変異誘発などの方法によって、新規の制限部位を作り出したり、グリコシル化パターンを変えたり、優先コドンを変えて特定の宿主における発現を増大させたり、スプライスバリエーションを作り出したり、半減期を延長する等の操作が可能である。この発現ベクターは、特定の宿主における各要素の効率に基づいて選択された様々なサンプルに由来する転写及び翻訳調節エレメント（プロモーター及びエンハンサー、特定の開始シグナル、ポリアデニル化3'配列

10

20

30

40

50

）を含み得る。in vitro組み換えDNA技術、合成技術及び/またはin vivo遺伝子組み換え技術を組み合わせて、このベクターにcDNAと調節エレメントをつなぐことができる。このような技術は、当分野で周知であり、Sambrook(前出、ch. 4, 8, 16 and 17)に記載されている。

【0061】

様々な宿主系を発現ベクターで形質転換することができる。以下に限定するものではないが、これらの中には組み換えバクテリオファージやプラスミド、またはコスミドDNA発現ベクターで形質転換された細菌と、酵母発現ベクターで形質転換された酵母と、バキュロウイルス発現ベクターで形質転換された昆虫細胞系と、ウイルスエレメント及び/または細菌エレメントを含む発現ベクターで形質転換された植物細胞系や動物細胞系が含まれる (Ausubel前出、unit 16)。例えば、アデノウイルス転写/翻訳複合体を哺乳動物細胞に用いることができる。配列をウイルスのゲノムのE1若しくはE3領域に結合させた後、この感染ウイルスを用いて形質転換させ、宿主細胞でタンパク質を発現させることができる。また、ラウス肉腫ウイルスエンハンサーやSV40、またはEBV系のベクターを用いてタンパク質を高発現させることができる。

10

【0062】

核酸配列のルーチンのクローニング及びサブクローニング、増殖は、多機能PBLUESCRIPTベクター(Stratagene, La Jolla CA)またはPSPORT1プラスミド(Life Technologies)を用いて行うことができる。核酸配列をこれらのベクターの多数のクローニング部位に導入すると、lacZ遺伝子が破壊され、形質転換された細菌を確認するための比色法によるスクリーニングが可能となる。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での転写及びジデオキシ法によるシーケンシング、ヘルパーファージによる一本鎖の調整、入れ子状欠失の作製において有用である。

20

【0063】

長期に渡って組み換えタンパク質を産生させるために、同一或いは別のベクター上の選択マーカー遺伝子或いは可視マーカー遺伝子と共にこのベクターを持続的に細胞株に形質転換することができる。形質転換後、細胞を強化培地で約1~2日間増殖させてから選択培地に移す。選択マーカー、代謝拮抗物質、抗生物質、または除草剤耐性遺伝子は、関連する選択薬に対する抵抗性を与え、導入配列を確実に発現する細胞の増殖及び回収が可能となる。可視マーカーの発現によって同定された、或いは選択培地に生存することによって同定された耐性クローンを、培養技術を用いて増殖することができる。また、可視マーカーを用いて、導入された遺伝子によって発現するタンパク質を定量することができる。宿主細胞が目的のcDNAを含むか否かの決定は、DNA-DNAまたはDNA-RNAハイブリダイゼーション、或いはPCR増幅技術に基づいて行うことができる。

30

【0064】

宿主細胞は、組み換えタンパク質を目的の形に修飾する能力に基づいて選択することができる。このような修飾には、アセチル化及びカルボキシル化、グリコシル化、リン酸化、脂質化、及びアシル化等が含まれる。「プレプロ」型を切断する翻訳後プロセッシングを利用して、タンパク質のターゲティング、折り畳み及び/または活性を特定することができる。翻訳後活性のための特定の細胞装置及び特徴的な機構を有するATCC(Manassas, MD)から得られる異なった宿主細胞が、組み換えタンパク質の適当な修飾及びプロセッシングが確実に行われるようにするために選択され得る。

40

【0065】

細胞培地からのタンパク質の回収

精製を容易にするために、ベクターに導入する異種部分は、グルタチオンSトランスフェラーゼ(GST)、6-His、FLAG、MYC等を含む。GST及びCBP、6-Hisはそれぞれ、グルタチオン及びカルモジュリン、金属キレート樹脂が結合した市販のアフィニティマトリックスを用いて精製される。FLAG及びMYCは、市販のモノクローナル抗体及びポリクローナル抗体を用いて精製される。精製に続く分離を容易にするために、タンパク質の分割部位をコードする配列は、タンパク質と異異形部分との間のベクターの一部であって良い。組み換

50

えタンパク質の発現及び精製の方法はAusubel(前出、unit 16)に記載され市販されている。

【0066】

ペプチドの化学合成

タンパク質若しくはその一部は、組み換え方法以外の当分野で周知の化学的方法によって合成することもできる。固相技術を用いるペプチド合成は、バッチ式或いは連続的なフロープロセスによって行うことができる。連続的なフロープロセスでは、アミノ保護及び側鎖保護アミノ酸残基をリンカーを介して不溶性の高分子支持物に連続的に追加する。メチルアミン誘導体化ポリエチレングリコールなどのリンカーを、ポリ(スチレン-co-ジビニルベンゼン)に結合させて支持レジンを形成する。このアミノ酸残基は、酸不安定Boc(*t*-butyloxycarbonyl)法若しくは塩基不安定Fmoc(9-fluorenylmethoxycarbonyl)法によって保護されたN-である。保護されたアミノ酸のカルボキシル基をリンカーのアミンに結合して、この残基を固相支持レジンに結合させる。Boc若しくはFmocを用いた場合、トリフルオロ酢酸若しくはピペリジンを用いて保護基を除去する。カップリング試薬若しくは予め活性化されたアミノ酸誘導体を用いて、追加する各アミノ酸を結合された残基に付加してから、レジンを洗浄する。完全長のペプチドは、連続的な保護の停止、即ち誘導体化アミノ酸を結合させて合成し、ジクロロメタン及び/またはN、N-ジメチルホルムアミドで洗浄する。このペプチドは、ペプチドカルボキシル末端とリンカーとの間で切断され、ペプチド酸またはペプチドアミドが作られる(Novabiochem 1997/98 Catalog and Peptide Synthesis Handbook, San Diego CA pp. S1-S20)。ABI 431 A ペプチドシンセサイザー(Applied Biosystems)などの装置を用いて、ペプチドを自動合成することができる。タンパク質またはその一部は調整用の高性能液体クロマトグラフィーによって精製し、その組成をアミノ酸解析またはシーケンシングによって確認することができる(Creighton (1984) *Proteins. Structures and Molecular Properties*, WH Freeman, New York NY)。

【0067】

抗体の準備及びスクリーニング

ヤギ、ウサギ、ラット、マウス、及びヒト等を含む様々な宿主は、STEAPRP若しくはその任意の一部を注入して免疫することができる。フロイントなどのアジュバント及びミネラルゲルと、リゾレシチン及びpluronic polyol、ポリアニオン、ペプチド、油乳剤、キーホールリンペットヘモシニアン(KLH)、ジニトロフェノールなどの表面活性物質とを用いて免疫反応を高めることができる。オリゴペプチドやペプチド、またはタンパク質の一部を用いて、少なくとも約5個のアミノ酸、より好ましくは10個の天然のタンパク質の一部と同一のアミノ酸を含む抗体を誘発させる。キメラ分子に対する抗体を産生させるために、オリゴヌクレオチドをKLHなどのタンパク質と融合させることができる。

【0068】

モノクローナル抗体は、培地の連続細胞株によって抗体を産生させる任意の技術を用いて準備する。以下に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及びヒトB細胞ハイブリドーマ技術、EBV-ハイブリドーマ技術が含まれる(例えば、Kohler 他(1975) *Nature* 256:495-497; Kozbor 他(1985) *J. Immunol. Methods* 81:31-42; Cote 他(1983) *Proc. Natl. Acad. Sci.* 80:2026-2030; and Cole 他(1984) *Mol. Cell Biol.* 62:109-120.を参照)。

【0069】

別法では、当分野で周知の方法を用いて、抗体生成のために記載された技術が適用され、エピトープ特異的一本鎖抗体を生産する。タンパク質のエピトープに対して特異的に結合する部位を含む抗体断片を生産することが可能である。限定するものではないが、このような断片には、例えば、抗体分子のペプシン消化によって作製されたF(ab')₂断片及びこのF(ab')₂断片のジスルフィド架橋を減少させて作製したFab断片が含まれる。別法では、Fab発現ライブラリを作製して、目的の特異性を有するモノクローナルFab断片の高速かつ容易に同定できるようにする(例えば、Huse 他(1989) *Science* 246:1275-1281を参照)。

【 0 0 7 0 】

STEAPRP若しくはその一部を用いて、ファージミドまたはBリンパ球免疫グロブリン・ライブラリをスクリーニングして、目的の特異性を有する抗体を同定する。確立された特異性を有するモノクローナル抗体或いはポリクローナル抗体のいずれか一方を用いる、競合的結合またはイムノアッセイの様々なプロトコルが当分野で周知である。このようなイムノアッセイは通常、このタンパク質とその特異的な抗体との複合体形成の測定を行う。2つの非干渉エピトープに反応するモノクローナル抗体を用いる2部位モノクローナル系イムノアッセイが好ましいが、競合的結合によるアッセイを用いることもできる (Pound (1998) Immunochemical Protocols. Humana Press, Totowa NJ)。

【 0 0 7 1 】

10

アッセイのための分子の標識化

多様なレポータ分子及び接合技術が当分野で周知であり、様々な核酸やアミノ酸、及び抗体のアッセイに用いることができる。標識した分子の合成は、 ^{32}P -dCTPまたはCy3-dCTP、Cy5-dCTP (Operon Technologies, Alameda CA) などの標識したヌクレオチドや ^{35}S メチオニン (APB) などのアミノ酸を組み込むための市販のキット (Promega, Madison Wis) を用いて行うことができる。ヌクレオチド及びアミノ酸は、BIODIPYまたはFITC (Molecular Probes, Eugene OR) などの試薬を用いて分子中に存在するアミン及びチオール基または他の基に化学的に結合させることで、様々な物質 (蛍光剤または化学発光剤、色素産生剤など) で直接標識することができる。

【 0 0 7 2 】

20

(診断)

本cDNA、断片、オリゴヌクレオチド、相補的なRNA及びDNA分子、及びPNAは、疾患の診断のために、遺伝子発現の変化を検出及び定量するのに用いられても良い。STEAPRPと特異的に結合する同様の抗体は、タンパク質を定量化するのに用いられてもよい。発現変動に関連する疾患には、とりわけ前立腺過形成もしくは前立腺癌のような前立腺細胞増殖性疾患が含まれる。診断アッセイにハイブリダイゼーションまたは増幅技術を用いて、遺伝子発現の変化を検出するべく、患者からの生体サンプルの遺伝子発現レベルを標準的なサンプルの値と比較する。質的または量的なこのような比較法は当分野で周知である。

【 0 0 7 3 】

例えば、本核酸分子またはプローブを標準的な方法で標識して、これを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、患者からの生体サンプルに加える。インキュベーションの後、このサンプルを洗浄し、ハイブリダイゼーション複合体に関連する標識 (シグナル) の量を定量して標準値と比較する。患者のサンプルに於けるコンプレクスフォーメーションが通常もしくは疾患標準値の何れかと著しく異なっている (高いまたは低い) 場合は、疾患の存在を示唆する。

【 0 0 7 4 】

差次的な発現を確立するための基準を設けるために、通常及び疾患発現プロファイルが確立される。この発現プロフィールは、ハイブリダイゼーションまたはに好適な条件下で、ヒト若しくは動物の正常被験体から採取したサンプルを、cDNAと結合させることによって確立される。標準的なハイブリダイゼーション複合体は、正常な被験体から得た値と、精製された配列を所定量用いた実験値とを比較することによって定量することができる。このように求めた標準値を、特定の症状や疾患、または異常症を示す患者のサンプルから得た値と比較することができる。標準値と特定の疾患に関連する値との偏差からその症状を診断する。

【 0 0 7 5 】

またこのようなアッセイを用いて、動物実験や臨床検査における特定の治療計画の効果を評価したり、患者個人の治療をモニタリングすることができる。病態が確認されると治療プロトコルを開始し、通常ベースで診断アッセイを繰り返して、被験者における発現のレベルが正常な患者に示される値に近づき始めたか否かを調べることが可能である。連続して行ったアッセイの結果から、数日から数ヶ月に渡る期間の治療効果を調べることができ

50

る。

【0076】

免疫学的方法

特異的なポリクローナル抗体若しくはモノクローナル抗体の何れかを用いるタンパク質の検出及び定量は当分野で周知である。このような技術には、ELISA（酵素結合免疫吸着検定法）及びラジオイムノアッセイ（RIA）、蛍光活性化セルソーター法（FACS）が含まれる。2つの非干渉エピトープに反応するモノクローナル抗体を用いる2部位モノクローナル系イムノアッセイが好ましいが、競合的結合アッセイを用いることもできる（例えば、Coligan 他（1997）Current Protocols in Immunology, Wiley-Interscience, New York NY; および Pound 前出）。

10

【0077】

（治療）

図2A、2B、及び2Cに示すように、STEAPRP（SEQ ID NO: 1）とヒトSTEAP（g6572948; SEQ ID NO: 11）との間の領域には、とりわけ6つの膜貫通ドメイン内に、化学的及び構造的類似性が存在する。加えて、発現変動は、表1乃至3に示されているように、LNCaP前立腺癌細胞及び前立腺組織に高度に関連し、とりわけとりわけ前立腺過形成もしくは前立腺癌のような前立腺細胞増殖性疾患に関連する。STEAPRPはとりわけ前立腺過形成もしくは前立腺癌のような前立腺細胞増殖性疾患に際だった役割を果たす。

【0078】

STEAPRPの増加発現に関する状況の治療では、発現若しくはタンパク質活性を減少させることは不可能である。ある実施例では、タンパク質の抑制遺伝子、アンタゴニスト、及び抗体が被験者に対して投与され、増加した発現及び活性に関するコンディションを治療する。別の実施例では薬剤キャリアに関連する抑制遺伝子、アンタゴニスト、及び抗体を含む薬剤成分が、被験者に対して投与され、内因性タンパク質の増加発現及び活性に関するコンディションを治療する。追加的な実施例では、cDNA若しくはその断片の補体を発現するベクターが、疾患の治療のために被験者に投与される。

20

【0079】

STEAPRPの減少した発現に関連する症状の治療では、発現若しくはタンパク質活性を増加させることが出来ない。ある実施例では、タンパク質、アゴニスト、若しくはエンハンサが、被験者に投与され、減少した発現若しくは活性に関する症状を治療する。別の実施例では、薬剤のキャリアに関連して、タンパク質、アゴニスト、及びエンハンサを有する薬剤成分が被験者に対して投与され、内因性タンパク質の減少発現若しくは活性に係る症状を治療する。さらなる実施例では、cDNAを発現するベクターが、疾患を治療する目的で被験者に投与される。

30

【0080】

本cDNA、またはそれに相補的な分子やその一部、そのタンパク質やその一部の内の任意のもの、これらの核酸分子やタンパク質を運ぶベクター、及びそれらのリガンドをその他の薬剤と共に投与することが可能である。併用療法に用いる薬剤の選択は、当業者が従来の薬学原理に従って行うことができる。薬剤を併用することによって、少量の各薬剤で特定の症状の予防または治療において相乗的な効果をあげることが可能である。

40

【0081】

核酸を用いる遺伝子発現の調節

遺伝子の発現は、STEAPRPをコードする遺伝子の5'または3'調節領域、或いは他の調節領域に対して相補的或いはアンチセンス分子を設計することで調節することが可能である。転写開始部位に対して設計されたオリゴヌクレオチドが好ましい。同様に、ポリメラーゼ、転写因子、または調節分子の結合を阻止する三重螺旋塩基対合で遺伝子発現を阻止することができる（Gee 他 In: Huber and Carr (1994) Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177）。また、リボソームとmRNAとの結合を阻止して翻訳が行われないように、相補的な分子を設計することも可能である。或るいは、複数のcDNAまたはその断片のライブラリをスクリーニングして、翻訳されない

50

調節配列に特異的に結合する核酸分子または断片を同定することも可能である。

【0082】

また、酵素活性をもつRNA分子であるリボザイムを用いて、RNAの特異的な切断を触媒してもよい。リボザイム作用のメカニズムは、まずリボザイム分子と相補的な標的RNAとの配列特異的なハイブリダイゼーションが起こり、次にGUAおよびGUU、GUCなどの部位においてヌクレオチド鎖が切断される。このような部位が一旦同定されたら、オリゴヌクレオチドを機能不全に示す二次構造特性について、同じ配列を有するオリゴヌクレオチドを評価することができる。また、候補標的としての適合性は、RNA分解酵素保護アッセイを用いて、相補的なオリゴヌクレオチドとのハイブリダイゼーションを検査して評価することができる。

10

【0083】

本発明の相補的な核酸およびリボザイムは、固相ホスホラミダイト化学合成法を用いて、*in vitro*または*in vivo*での組換え発現によって調製することが可能である。更に、RNA分子は、その5'および/または3'末端に隣接配列を付加して、或いは分子のバックボーン内のホスホジエステル結合の代わりにホスホロチオネートまたは2'-O-メチルを用いて、細胞内の安定性および半減期が増大するように改変することができる。この改変はPNAの作製に固有であるが、他の核酸分子にも適用することができる。例えばイノシン、queosine、wybutosineなどの伝統的でない塩基を含めて、またはアセチル基、メチル基、チオ基で、ウリジン、アデニン、シチジン、グアニン、およびチミンを修飾して、内在性エンドヌクレアーゼに対する分子の有効性を低くする。

20

【0084】

スクリーニング及び精製のアッセイ

本STEAPRPをコードするcDNAを用いて、分子若しくは混合物のライブラリをスクリーニングして特異的な結合親和性を調べることが可能である。このライブラリは、内在性遺伝子の活性、複製、転写、または翻訳を調節するアプタマー、DNA分子、RNA分子、PNA、ペプチド、転写因子などのタンパク質、エンハンサー、リプレッサー、およびその他のリガンドを含み得る。このアッセイは、特異的な結合を可能とする条件下でポリヌクレオチドと分子ライブラリとを結合させる課程と、特異的結合を検出して、1本鎖もしくは2本鎖の分子に特異的に結合する少なくとも一つの分子を同定する課程とを含む。

30

【0085】

一実施例において、本発明のcDNAを複数の精製された分子または複合物と共にインキュベートすることが可能であり、電気泳動度移動アッセイ(USPN 6,010,849)または網状赤血球溶解物転写アッセイ等の当分野で公知である方法によって結合活性を測定することが可能である。別の実施例において、cDNAは生検そして/あるいは培養細胞と組織からの核抽出物と共にインキュベートすることが可能である。核抽出物のcDNAと分子あるいは化合物の間の特異結合は、最初ゲルシフトアッセイにより決定される。そして後にその分子または化合物を回収して、それに対する抗体を作製することによって確認することが可能である。これらの抗体がアッセイに加えられる時、電気泳動度移動アッセイにスーパーシフトが起きる。

40

【0086】

別の実施例では、cDNAは、当分野で公知であるアフィニティークロマトグラフィー方法を用いて分子あるいは化合物を精製するために使用される。一実施例では、cDNAは高分子樹脂またはゲル上で臭化シアン基と化学的に反応する。次にサンプルを通し、cDNAと反応あるいは結合させる。cDNAに結合する分子または化合物は、流動媒体の塩の濃度の上昇によりcDNAから放出され、回収される。

【0087】

更なる実施例において、タンパク質あるいはその部分をサンプルのリガンドを精製するために用いる可能性がある。リガンドを精製するためにタンパク質またはその部分を使用する方法には、特異結合を許容する条件下でタンパク質またはその部分をサンプルと結合すること、タンパク質とリガンド間の特異結合を検出すること、結合したタンパク質を回収

50

すること、精製したリガンドからタンパク質を分離するために適切なカオトロピック剤を使用することが含まれる。

【0088】

好適な実施例では、STEAPRPを多様なスクリーニングアッセイの任意の複数の分子または化合物をスクリーニングするために用い得る。そのようなスクリーニングで用いられたタンパク質の部分は、溶液中で遊離しているか、非生物的物质または生物的基質に固定させる（例えば細胞表面上に保持される）、あるいは細胞内に位置することになる。例えば1つの方法では、組換え核酸で安定的に形質転換された、生きたまたは固定された原核宿主細胞で、細胞表面にペプチドを発現し局在させる細胞をスクリーニングアッセイで使用する事が可能である。細胞を複数のリガンドあるいはリガンドのライブラリに対してスクリーニングする。そして発現したタンパク質とリガンドの間で結合の特異性または複合体の形成を測定し得る。タンパク質と分子間の特異結合を測定してもよい。スクリーニングされたライブラリの種類に応じて、アッセイをDNA分子、RNA分子、ペプチド核酸、ペプチド、タンパク質、擬態、アゴニスト、アンタゴニスト、抗体、免疫グロブリン、インヒビター、薬剤及び他のリガンドを同定するために用いる可能性があり、そしてタンパク質を特異結合する。

10

【0089】

或る実施態様において、本発明はUSPN 5,876,946に記載された非常に小さなアッセイ量と微量を使用して試験化合物高い処理能力でスクリーニングするための方法を網羅する。当該文献は特に引用することを以って本明細書の一部となす。この方法は、特異結合により多数の分子と化合物をスクリーニングするために用いられる。別の実施態様において本発明は、タンパク質を結合することができる中和抗体が、タンパク質を結合するための試験化合物と特異的に競合する、競合的薬剤スクリーニングアッセイの使用も網羅する。スクリーニングによって同定された分子または化合物は、毒性、診断または可能性のある治療薬を評価するために哺乳動物モデル系で用いられ得る。

20

【0090】

薬理学

医薬組成物とは、所望の目的を達成するのに効果的な量の活性成分を含んでいる物質である。効果的な薬用量の決定は、当分野の技術者の能力による部分が多い。どんな化合物であっても、初めは細胞培養アッセイ或いは動物モデルの何れかによって治療効果のある薬用量を推定する。また、動物モデルを使って、好適な濃度範囲および投与経路を決定する。次に、このような情報を用いて、ヒトへの効果的な投与経路および薬用量を決定する。

30

【0091】

治療効果のある薬用量とは、症状または病態を改善するタンパク質またはインヒビターの量である。このような薬剤の薬用効果および毒性は、例えば、ED₅₀（集団の50%に医薬的效果がある投与量）およびLD₅₀（集団の50%に致命的である投与量）などの細胞培養または実験動物における標準的な製薬方法によって決定することができる。或る投与量における毒性効果と治療効果との比率が治療指数となり、LD₅₀ / ED₅₀と示すことができる。高い治療指数を示す医薬組成物が好ましい。細胞培養アッセイおよび動物実験から得られたデータを用いて、ヒトへ適用する薬用量の範囲を決定する。

40

【0092】

（モデル系）

動物モデル系は生物学的検定法として用い得る。そこではヒトと類似の表現型反応を示し、また暴露条件が人体暴露と関連する哺乳動物は最も一般的なモデルであり、多くの感染体、癌、薬剤および毒性研究が、低費用、利便性、寿命、生殖可能性、豊富な参考文献ゆえにラットまたはマウス等の齧歯類で実施される。近交系また非近交系齧歯類は、目的遺伝子の過小発現と過剰発現の生理作用の結果の研究、および疾患の診断と治療のための方法開発のための便利なモデルを提供する。特定の遺伝子を過剰に発現する哺乳動物近交系（例えば乳汁内に分泌する）は、その遺伝子により発現する便利なタンパク質源となり得

50

る。

【0093】

毒性研究

毒性研究は、生物系上での薬剤影響の研究である。多くの毒性研究は、ラットまたはマウスで実施される。生理学、行動、恒常性プロセスにおける定性または定量変更そしてラットまたはマウスの死亡率の観察は、毒性プロフィールを生み出し、薬剤への暴露の後にヒトの健康に及ぼす潜在的な因果関係を評価するために使われる。

【0094】

遺伝毒性研究により、内在性、自発的そして誘導された遺伝突然変異の率への薬剤の影響の同定と分析がなされる。遺伝毒性薬剤は、核酸との相互作用を促進する共通の化学的特性また物理的特性を通常は有する。また染色体の異常が子孫に伝達された時最も有害である。毒性研究により、受胎前に親に、妊娠中に母親にあるいは発達中の生物のいずれかに投与された場合子孫の組織中の構造的また機能的異常の頻度が増大する物質を同定することが可能である。マウスまたはラットは、これらの試験で最も頻繁に使用される。なぜなら生殖周期が短いので統計上の必要にかなう多数の生物の繁殖を可能にするからである。

10

【0095】

急性毒性試験は、症状改善または薬剤の致死率を決定するために被験動物に薬剤を1回投与することに基づいている。3回の実験が実施される。：1)初回投与量範囲を定める実験、2)有効量の範囲を狭める実験、3)用量反応曲線を確定するための最終実験。

【0096】

亜慢性毒性試験は、薬剤の連続投与に基づく。ラットとイヌはこれらの研究で一般的に用いられ、様々な属の種からのデータを提供する。発癌を除いて、3-4ヶ月の間、高投与量濃度で薬剤を毎日投与することにより、成獣における毒性の多くの形態を明らかにできることにに関して相当の証拠がある。

20

【0097】

1年以上の期間の慢性毒性試験により、毒性の不存在か薬剤による発癌の可能性を実証する。ラットによって研究が実行される時、最小限の3つの試験グループに加えて1つの対照グループが用いられる。実験当初および実験中は定期的に動物は調べられ、モニターされる。

【0098】

遺伝子組換え動物モデル

目的の遺伝子を過剰発現あるいは過小発現する遺伝子組換え齧歯類は近親交配され、ヒト疾患のモデル、または治療薬や毒性薬剤の試験のために用いられ得る。(USPN 5,175,383、USPN 5,767,337等を参照.)。場合によっては、導入された遺伝子は胎児の発育中または出生後の発育中の特定の時間に特定の組織のタイプで活性化され得る。導入遺伝子の発現は、薬物治療法実験への取り組み前、中、後に遺伝子組換え動物の表現型、組織特異的mRNA発現または血清と組織タンパク質レベルの分析によってモニターされる。

30

【0099】

胚性幹細胞

齧歯類の胚から単離された胚性幹細胞(ES細胞)は、胚組織を形成する可能性を保持する。ES細胞がキャリアの胚内部に置かれた時、正常な発育を再開して、生きて生まれた動物の組織に寄与する。ES細胞は実験的ノックアウトとノックイン齧歯類の生成に用いられる好適な細胞である。129/SvJ細胞株等のマウスES細胞は初期のマウス胚に由来し、当分野で公知の培養条件下で増殖させることができる。遺伝子組換え類を産生するのに用いられたベクターには、疾患遺伝子候補と標識遺伝子が含まれる。後者は導入された疾患遺伝子の存在を同定することに役立つ。ベクターは当分野で公知の方法でES細胞に形質転換する。そして形質転換したES細胞を同定し、例えばC57BL/6マウス系等から採取したマウス細胞胚盤胞に微量注入する。胚盤胞を偽妊娠メスに外科的に導入し、得られるキメラ子孫の遺伝形質を決め、これを交配させてヘテロ接合性系またはホモ接合性系を作製する。

40

【0100】

50

ヒト胚盤胞由来のES細胞は、invitroで少なくとも8つの別々の細胞系統に分化するよう操作することが可能である。これらの系統はinvitroで様々な細胞のタイプと組織の分化を研究するのに用いられ、例えば神経細胞、造血系統及び心筋細胞に分化する内胚葉、中胚葉及び外胚葉の細胞の種類を含む。

【0101】

ノックアウト分析

遺伝子ノックアウト分析では、哺乳動物遺伝子の領域はネオマイシンホスホトランスフェラーゼ遺伝子 (neo: Capecchi(1989) Science 244:1288-1292) 等の非哺乳動物遺伝子を含むよう酵素処理で修飾される。修飾された遺伝子は、培養したES細胞に形質転換され、相同組換えにより内因性ゲノムに組み込まれる。挿入された配列は、内在性遺伝子の転写と翻訳を破壊する。形質転換細胞を齧歯類胚に注入し、その胚を偽妊娠メスに移植する。遺伝子組換え子孫は、哺乳動物遺伝子の機能複製を欠くホモ接合性近交系を得るために交配される。一例では、哺乳動物遺伝子はヒト遺伝子である。

10

【0102】

ノックイン分析

ES細胞を用いて、ヒト疾患の「ノックイン」ヒト化動物(ブタ)または遺伝子組換え動物モデル(マウスまたはラット)を作製することが可能である。ノックイン技術を用いて、ヒト遺伝子の或る領域を動物ES細胞に注入し、注入したヒト配列を動物細胞ゲノムに組み込ませる。形質転換細胞を胚に注入し、胚を上記のように移植する。遺伝子組換え子孫または近交系について研究し、可能性のある医薬品を用いて処理し、類似のヒトの症状の治療に関する情報を得る。これらの方法は、幾つかのヒト疾患をモデル化するために用いられてきた。

20

【0103】

ヒト以外の霊長類モデル

動物実験の分野では、生理学、遺伝学、化学、薬理学、統計学等の基礎科学のデータと方法論を扱う。これらのデータは、ヒトの健康に関連している可能性があるので、ヒト以外の霊長類での治療薬の影響を評価するのに非常に重要である。ワクチンと薬物の評価においてサルがヒトの代用に用いられる。またサルの反応が類似の条件下で人体暴露と関連される。このような調査では、カニクイザルとアカゲザル(それぞれ Macaca fascicularis と Macaca mulatta)、コモンマモセット (Callithrix jacchus) が最も一般的に用いられるヒト以外の霊長類(NHP)である。NHPのコロニーの発達と維持には多額の費用がかかるため、齧歯類のモデルで初期の調査と毒性研究が通常は実施される。薬物常用などの行動測定を利用する研究では、NHPが最初に実験動物に選ばれる。さらに、NHPと個々のヒトは多くの薬剤と毒素に異なる感受性を示す。またこれらの薬剤の「広範囲のメタボライザー」から「代謝不良体質」まで表現型の範囲として分類することが可能性である。

30

【0104】

別の実施例では、将来に開発される分子生物学技術で、現在知られているヌクレオチド配列の特性(限定はされないが、トリプレット遺伝コード、特異的な塩基対相互作用等を含む)に依存しているならば、このタンパク質をコードするcDNAにその新技術を用い得る。

【実施例】

40

【0105】

下記の実施例は、目的の本発明を説明するために提供するが、これは本発明を限定する目的ではない。ヒト脳鋸歯状核BRAWTDR02ライブラリの準備について記載する。

【0106】

1 cDNAライブラリの作製

BRAWTDR02 cDNAライブラリは、胆管癌で死亡した55歳の白人女性より採取した脳鋸歯状核組織(標本#A98-58)で作成された。凍結組織はPOLYTRONホモジナイザー(Brinkmann Instruments, Westbury NJ)を用いてTRIZOL 試薬(0.8 g tissue/12 ml; Life Technologies)内でホモジナイズ及び溶解された。ライセートは、室温・25,000 rpm・18時間に渡り、L8-70M 超遠心機(Beckman Coulter, Fullerton CA)内のSW28ローターを用い、5.7M CsC

50

1クッションで遠心分離された。RNAは、4.7の酸フェノールで抽出され、0.3 Mの酢酸ナトリウム及び2.5 倍のエタノールを用いて沈殿され、RNAseを含まない水中で再懸濁され、また37 でDNAse処理される。OLIGOTEX精製キット (QIAGEN, Chatsworth CA) を用いて、RNAが単離され、cDNAライブラリの構築に用いられた。

【0107】

これをcDNAライブラリの作製に用いた。このmRNAを、mRNAのポリ(A)尾部で一本鎖cDNAの合成を開始するように設計されたNotIプライマー - アダプターを含むSUPERScript プラスミドシステム (Life Technologies) の推奨プロトコルに従って処理した。二本鎖cDNAを平滑化し、EcoRIアダプターに結合し、NotI (New England Biolabs, Beverly MA) で消化した。このcDNAを、SEPHAROSE CL4Bカラム (APB) 上で分画化し、400bpを超える大きさのcDNAを、pINCYプラスミド (Incyte Genomics) に結合させた。このプラスミドpINCYは、DH5 コンピテント細胞 (Life Technologies) に形質転換された。

10

【0108】

2 cDNAクローンの単離とシーケンシング

プラスミドDNAを細胞から放出して、MINIPREPキット (Edge Biosystems, Gaithersburg MD) またはREAL PREP 96 プラスミドキット (Qiagen) のいずれかを用いて精製した。このキットには、960精製のためリガンドでブロックする96-ウェルから成る。推奨されているプロトコルを下記の変更を除いて使用する。1) 細菌を、カルペニシリン25 mg/lと0.4%のグリセロールの入った、殺菌したTERRIFIC BROTH (BD Biosciences, Sparks MD) 1 mlで培養した。2) 接種後、細胞を19時間培養し、次いで0.3 mlの溶解バッファで溶解した。3) イソプロパノールによる沈殿の後、プラスミドDNAペレットを0.1 mlの蒸留水で再懸濁した。プロトコルの最終ステップ後に、サンプルを96-ウェルブロックに移して、4 で保管した。

20

【0109】

cDNAをDNA ENGINE サーマルサイクラー (MJ Research) と併用してMICROLAB 2200システム (Hamilton) を使用して、シーケンシングのために調製した。cDNAをABI PRISM 377シーケンシングシステム (Applied Biosystems) またはMEGABACE 1000 DNAシーケンシングシステム (APB) を用いてSangerとCoulson (1975; J Mol Biol 94:441-448) の方法で配列化した。多くの単離は、溶液体積0.25x-1.0x濃度の標準ABIプロトコルとキット (Applied Biosystems) に従ってシーケンスされた。別法では、cDNAをAPBの溶液と色素を用いてシーケンスした。

30

【0110】

3 cDNA配列の伸長

cDNAクローンとオリゴヌクレオチドプライマーを用いて、cDNAを伸長した。一方のプライマーは既知の断片の5'伸長を開始するべく合成し、他方のプライマーは既知の断片の3'伸長を開始するべく合成した。開始プライマーは、市販のソフトウェア (Molecular Biology Insights) を用いて、約22個から約30個のヌクレオチドの長さで約50%以上のGC含量を有し、かつ約68~72 の温度で標的配列にアニールするように設計した。ヘアピン構造及びプライマー - プライマー二量体が生じないようにヌクレオチドを伸長した。

40

【0111】

配列を伸長するために、鋳型として選択されたcDNAライブラリを用いた。2段階以上の伸長が必要な場合には、付加的プライマー或いはプライマーのネステッドセットを設計した。好適なライブラリのサイズを選択して、更に遺伝子の5'または上流領域を有する配列を含めるために大きなcDNAとランダムプライマーを含めた。ゲノムライブラリを特に5'プロモーター結合領域への伸長する調節領域を得るために使用する。

【0112】

高忠実度の増幅をUSPN 5,932,451で開示された方法を利用したPCR法によって得た。PCR法をDNA ENGINE サーマルサイクラー (MJ Research.) を用いて96穴プレート内で実施した。反応混合液は、鋳型DNA及び200 nmolの各プライマー、 Mg^{2+} と $(NH_4)_2SO_4$ と - メ

50

ルカプトエタノールを含む反応バッファ、Taq DNAポリメラーゼ (APB)、ELONGASE酵素 (Life Technologies)、Pfu DNAポリメラーゼ (Stratagene) を含む。プライマーの組、PCI AとPCI B (Incyte Genomics) に対して以下のパラメータで増幅を行った。: ステップ 1: 94 で3分間、ステップ 2: 94Cで15秒間、ステップ 3: 60Cで1分間、ステップ 4: 68C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: 68Cで5分間、ステップ 7: 4 で保管。別法では、プライマーの組、T7とSK+ (Stratagene) に対して以下のパラメータで増幅を行った。ステップ 1: 94 で3分間、ステップ 2: 94Cで15秒間、ステップ 3: 57Cで1分間、ステップ 4: 68C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: 68Cで5分間、ステップ 7: 4 で保管。

【0113】

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

【0114】

伸長させたクローンは、脱塩及び濃縮して384穴プレートに移し、CviJIコレラウイルスエンドヌクレアーゼ (Molecular Biology Research, Madison WI) を用いて消化し、pUC 18ベクター (APB) への再連結反応前に音波処理またはせん断した。ショットガン・シーケンシングのために、消化したヌクレオチド配列を低濃度 (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AGARACE酵素 (Promega) で消化した。伸長させたクローンをT4DNA (New England Biolabs) を用いてpUC 18ベクター (APB) に再連結し、Pfu DNAポリメラーゼ (Stratagene) で処理して制限部位の張出部 (overhang) を満たし、大腸菌細胞に形質移入した。形質移入した細胞を選択して抗生物質を含む培地に移し、それぞれのコロニーを切りとってLB/2Xカルベニシリン培養液の384ウェルプレートに37 で一晚培養した。

【0115】

細胞を溶解して、Taq DNAポリメラーゼ (APB) 及びPfu DNAポリメラーゼ (Stratagene) を用いて以下の手順でDNAをPCR増幅した。ステップ 1: 94 で3分間、ステップ 2: 94 で15秒間、ステップ 3: 60 で1分間、ステップ 4: 72 で2分間、ステップ 5: ステップ 2、3 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 で保管。上記したようにPICOGREEN量的試薬 (Molecular Probes) でDNAを定量化した。DNAの回収率が低いサンプルは、上記と同一の条件を用いて再増幅した。サンプルは20%ジメチルスルホキシド (DMSO; 1:2, v/v) で希釈し、DYENAMIC エネルギートランスファー シークエンシングプライマー、及びDYENAMIC DIRECTサイクルシークエンシングキット (APB) またはABI PRISM BIGDYE ターミネーターサイクル シークエンシング反応キット (Applied Biosystems) を用いてシークエンシングした。

【0116】

4 cDNAクローンの相同性検索と推定タンパク質

配列表のcDNAまたは推定アミノ酸配列を用いて、GenBank、SwissProt、BLOCKSなどのデータベースに問合せした。BLASTかBLAST2を用いて前もって同定した、またアノテーションが付けられた配列あるいはドメインを含むこれらのデータベースを検索して、アラインメントを生成し、どの配列が厳密に一致するか、または相同的かを決定した。アラインメントを原核 (細菌性) または真核 (動物、カビあるいは植物) 生物のシークエンシングした。別法として、SmithとSmith (1992, Protein Engineering 5:35-51) で説明されているようなアルゴリズムを用いて、一次配列パターンと二次構造ギャップペナルティで処理することが可能であった。この出願で開示されているすべての配列は、少なくとも49ヌクレオチドの長さを有し、不要な塩基は1.2%に過ぎない (A、C、GまたはTよりもNが記録されて

10

20

30

40

50

いる)。

【0117】

前出のKarlinで詳説されているように、問合せ配列とデータベース配列の間のBLAST一致を統計的に評価して、ヌクレオチドに 10^{-25} 、ペプチドに 10^{-25} の閾値を満たす時のみ報告した。下記のように計算して積スコアによって、相同性も評価した。BLASTのヌクレオチドあるいはアミノ酸の一致率[問い合わせ配列と参照配列間]に%最大限BLASTスコアを乗じ[問い合わせ配列と参照配列の長さに基づく]、次いで100で除する。実験室で用いるハイブリダイゼーション法と比較して、厳密な一致の電子ストリンジェンシーが、約40の下限(不要な塩基のために1-2%のエラーがある)から70の100%一致まで設定された。

【0118】

BLASTソフトウェア式(NCBI, Bethesda MD, <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>)には、核酸分子をアラインメントする「blastn」、ヌクレオチド又はアミノ酸配列を対で直接比較するために用いるBLAST 2など多種の配列分析プログラムが含まれる。BLASTプログラムは、一般的には、ギャップ及びデフォルト設定に設定された他のパラメータと共に用いる。例えば: 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; 及び Filter: onである。同一性は配列全体の長さにより測定する。Brennerら(1998; Proc Natl Acad Sci 95:6073-6078, 文献は、引用することをもって本明細書の一部とする)は配列同一性によって構造的相同性を同定する能力に関してBLASTを分析した。少なくとも150残基の配列アラインメントには信頼できる閾値は30%の同一性であり、少なくとも70残基のアラインメントに関しては40%であることがわかった。

【0119】

この出願のcDNAをアセンブルしたコンセンサス配列あるいはLIFESEQ GOLDデータベースで見つけたテンプレートと比較した(インサイトゲノミクス)。cDNA、伸長、全長およびショットガン・シーケンシングプロジェクトの構成エレメント配列をPHRED分析にかけて、クオリティスコアを指定した。クオリティスコアが許容できるすべての配列を多様なブリプロセッシングとエディティングの経路にかけ、低質の3'末端、ベクターとリンカー配列、ポリAテイル、Alu繰返し、ミトコンドリア及びリボゾーム配列、細菌汚染配列を除去する。編集した配列は少なくとも長さが50 bpであり、ジヌクレオチド反復、Alu反復などの情報に乏しい配列と反復エレメントを"Ns"あるいはマスクしたものに置換した。

【0120】

編集した配列は、構築処理にかけられ、配列が遺伝子ピンに割り当てられた。各配列は1つのピンにのみ所属でき、各ピンの配列を構築して、鋳型を作製した。BLASTとCROSSMATCHを用いて、新たにシーケンスした部分を既存のピンに加えた。前記の部分配列をピンに加えるために、150以上のBLAST質のスコアと少なくとも82%の局所同一性のアラインメントを有しなければならない。PHRAPを用いて、各ピンの配列を構築した。DEEP PHRAPを用いて、幾つかの重畳部分配列を有するピンを構築した。部分配列の数と配向に基づいて、各鋳型の配向を決定した。

【0121】

ピンを互いに比較して、少なくとも82%の局所類似性を有するピンを結合して、再構築した。95%以下の局所同一性である鋳型を有するピンを分離した。STITCHER/EXON MAPPERアルゴリズムによって鋳型を分析にかけ、スプライス変異体、或いはスプライスエキソン、スプライス接合部、組織のタイプや疾患状態全体の選択的スプライシングされた遺伝子の示差発現などの存在の確率を分析した。構築処理を周期的に繰り返した。そしてBLASTを用いてGbpriなどのGenBankデータベースに対して鋳型にアノテーションを付けた。厳密な一致について、200の塩基対に対して95%の局所同一性から100の塩基対に対して100%の局所同一性を有すると定義した。また相同一致については、E-value (または確率値) $< 1 \times 10^{-8}$ を有すると定義した。鋳型をGENPEPTに対するフレームシフト型FASTxにもかけた。相同一致をE-value $< 1 \times 10^{-8}$ を有すると定義した。鋳型分析と構築については1999年3月25日に提出したUSSN 09/276,534に記載されている。

10

20

30

40

50

【0122】

構築に続いて、鋳型をBLAST、モチーフその他の機能的分析にかけて、1997年3月6日に提出したUSSN 08/812,290、USSN 08/811,758、1997年10月9日に提出したUSSN 08/947,845、1998年3月4日に提出したUSSN 09/034,807等に記載されているタンパク質階層に分類した。3つの全フォワードリーディングフレームで各鋳型を翻訳することにより、またHMMERソフトウェアパッケージ(Washington University School of Medicine, St. Louis MO; <http://pfam.wustl.edu/>)を用いて隠れマルコフモデル(HMM)を基にした保存されたタンパク質ファミリードメインのPFAMデータベースに対する各翻訳を検索することにより鋳型を分析した。MACDNASIS PROソフトウェア(Hitachi Software Engineering)とLASERGENEソフトウェア(DNASTAR)を用いてcDNAを更に分析した。そして公共のデータベース、例えばGenBankのげっ歯類、哺乳動物、脊椎動物、原核生物、真核生物のデータベースと、SwissProt、BLOCKS、PRINTS、PFAMそしてPrositeに問い合わせた。

10

【0123】

5 染色体マッピング

スタンフォード・ヒトゲノムセンター(SHGC)、ホワイトヘッド・ゲノム研究所(WIGR)、Genethon等の公的な情報源から入手可能な放射線ハイブリッド及び遺伝地図データを用いて、配列表に存在するcDNAが前もってマッピングされたかを測定した。マッピングされたKSRCCをコードするcDNAの断片は、すべての関連する調節配列とコード配列を同じ位置に割り当てる。遺伝地図上の位置は、ヒト染色体の範囲または間隔として表される。cM間隔の地図上の位置は、染色体のpアームの末端に関連して測定する(ヒト中のDNAの1メガベース(Mb)にほぼ等しい)。

20

【0124】

6 ハイブリダイゼーション技術と分析基質上でのcDNAの固定

下記の方法の1つによってcDNAを基質に適用する。ゲル電気泳動法によりcDNAの混合を分画し、キャピラリー転移によりナイロン膜に転移する。別法として、cDNAを個々にベクターに連結反応させ、細菌性宿主細胞に挿入してライブラリを形成する。次に下記の方法の1つによってcDNAを基質に配置する。最初の方法では、個々のコロニーを含む細菌細胞をロボット利用して切りとり、ナイロン膜に配置する。膜を選択培地(用いられたベクターに依存してカルペニシリン、カナマイシン、アンピシリンまたはクロラムフェニコール)を含むLB寒天に置き、37℃で16時間インキュベートする。寒天から膜を取り除き、コロニーを上側にして、10% SDS、変性溶液(1.5 M NaCl, 0.5 M NaOH)、中和溶液(1.5 M NaCl, 1 M Tris, pH 8.0)、2xSSC(2回)に各10分、連続的に膜を置く。次に膜をSTRATALINKER UV架橋剤(Stratagene)でUV照射する。

30

【0125】

2番目の方法では、インサートに隣接するベクター配列に相補的なプライマーを用いて、cDNAを30サイクルのPCRで細菌ベクターから増幅する。PCR増幅で1~2 ngの初期量の核酸から5 µgより大きい最終量まで増大する。SEPHACRYL-400 ビーズ(APB)を用いて長さが400 bpから約5000 bpまで増幅した核酸を精製する。手であるいはドット/スロットブロット法マニホールドと吸気装置を用いて精製核酸をナイロン膜に置く。そして上述した変性、中和、UV照射により固定する。USPN 5,807,522に記載されている方法を用いて、精製した核酸を、ロボットで配置して、ポリマーコートされたスライドガラス上に固定する。顕微鏡スライドガラス(Corning, Acton MA)を良く洗って0.1%のSDS中で超音波をかけ、4%フッ化水素酸(VWR Scientific Products, West Chester PA)中でエッチングし、95%エタノール中で0.05%アミノプロピルシラン(Sigma Aldrich)を用いてコーティングする、そして110℃のオーブンで硬化させることにより、ポリマーコートされたスライドガラスを準備する。スライドガラスを処理前後で蒸留水中で広範囲にわたって洗浄する。核酸をスライドガラス上に置き、次にSTRATALINKER UV架橋剤(Stratagene)を用いてアレイをUV照射に暴露することにより固定する。アレイを室温において0.2% SDSで洗浄し、蒸留水で3度洗浄する。リン酸緩衝生理食塩水(PBS)(Tropix, Inc., Bedford

40

50

MA) 中の 0.2 % カゼイン中において 60 で 30 分間アレイをインキュベートした後、0.2 % SDS 及び蒸留水で洗浄することにより、非特異結合部位をブロックする。

【0126】

膜ハイブリダイゼーションのプロープ調製

配列表の cDNA に由来するハイブリダイゼーションのプロープを使用して、膜系ハイブリダイゼーションで cDNA、mRNA またはゲノム DNA をスクリーニングする。cDNA を 45 μ l TE バッファの濃度 40-50 ng に希釈し、5 分間 100 に加熱して変性させ、短時間遠心分離して、プロープを調製する。次に変性した cDNA を REDIPRIME 試験管 (APB) を加え、青色が均一に分布するまで軽く混合して、次に短時間遠心分離した。5 μ l の [32 P]dCTP をその試験管に加え、内容物を 37 で 10 分間インキュベートする。5 μ l の 0.2M EDTA を加えて標識化反応を停止する。そして PROBEQUANT G-50 microcolumn (APB) を用いてプロープを組み込まれていないヌクレオチドから精製する。精製したプロープを 5 分間 100 に加熱する、そして 2 分間氷上で急速に冷却して、下記に記載した膜系ハイブリダイゼーションで使用する。

10

【0127】

ポリマーコートされたスライドハイブリダイゼーションのプロープ調製

サンプルから分離した mRNA に由来するハイブリダイゼーションのプロープを使用して、アレイベースのハイブリダイゼーションで配列表の cDNA をスクリーニングする。9 μ l TE バッファで濃度 200 ng に mRNA を希釈し、5 μ l 5x バッファ、1 μ l 0.1 M DTT、3 μ l Cy3 または Cy5 標識化混合液、1 μ l RNアーゼ阻害因子、1 μ l 逆転写酵素、5 μ l 1x 酵母調節 mRNA を加え、GEMbright キット (Incyte Genomics) を用いてプロープを調製する。非コード酵母ゲノム DNA から *in vitro* 転写により酵母調節 mRNA を合成する (W. Lei, 未発表)。定量用対照として、サンプル mRNA に 0.002 ng、0.02 ng、0.2 ng、2 ng での対照 mRNA の 1 つの設定をそれぞれ 1:100,000、1:10,000、1:1000、1:100 (w/w) の比で逆転写反応混合液に希釈する。mRNA 示差発現パターンを調べるために、対照 mRNA の 2 番目の設定を 1:3、3:1、1:10、10:1、1:25、また 25:1 (w/w) の比で逆転写反応混合液に希釈する。反応混合液を混合して、37 で 2 時間インキュベートする。次に反応混合液を 20 分間、85 でインキュベートする。そして 2 つの連続する CHROMA SPIN+TE 30 カラム (Clontech, Palo Alto CA) を利用してプロープを精製する。精製したプロープは、プロープを DEPC 処理水の 90 μ l に希釈させ、2 μ l 1mg/ml グリコーゲン、60 μ l 5 M 酢酸ナトリウム、300 μ l 100% エタノールを加えて、エタノール析出させた。プロープを 20 分間、20,800xg で遠心分離する。そしてペレットを 12 μ l の再懸濁バッファで再懸濁し、5 分間 65 に加熱してから、十分に混合する。プロープを加熱して、以前のように混合して、氷上に保管する。以下に詳述するように、プロープを高密度アレイベースのハイブリダイゼーションにおいて使用する。

20

30

【0128】

膜系ハイブリダイゼーション

1% Sarkosyl と 1x 高リン酸バッファ (0.5 M NaCl, 0.1 M Na₂HPO₄, 5 mM EDTA, pH 7) を含むハイブリダイゼーション溶液で、55 で 2 時間、膜を前もってハイブリダイズする。15 ml の新鮮なハイブリダイゼーション溶液中で希釈されたプロープを次に膜に加える。膜をプロープと共に 55 で、16 時間、ハイブリダイズする。ハイブリダイゼーション後、1mM Tris (pH 8.0)、1% Sarkosyl で 15 分間、25 で膜を洗浄し、さらに 1mM Tris (pH 8.0) で、各回 15 分間、25 で 4 回洗浄する。ハイブリダイゼーション化合物を検出するために、XOMAT-AR フィルム (Eastman Kodak, Rochester NY) を膜に一晚 70 で暴露する。そして現像してから、視覚で確認する。

40

【0129】

ポリマーコートされたスライドハイブリダイゼーション

プロープを 5 分間 65 に加熱してから、5415C マイクロ遠心分離で 5 分間、9400 rpm で遠心分離する (Eppendorf Scientific, Westbury NY)。そして 18 μ l のアリクォットをアレイ表面に置き、カバーガラスで覆う。アレイは、顕微鏡用スライドより僅かに大きい空洞を

50

有する防水チェンバーに移す。チェンバーのコーナーに140 μ lの5 × SSCを加えることにより、チェンバー内部を湿度100%に保持する。このアレイを含むチャンバーを、60で約6.5時間インキュベートする。アッセイを1×SSC, 0.1% SDS中で、45で10分間洗浄し、次に0.1×SSCで、45で10分間の洗浄を3回繰り返した後に、乾燥させた。

【0130】

ハイブリダイゼーション反応を絶対的または示差ハイブリダイゼーションフォーマットで実施する。絶対ハイブリダイゼーションフォーマットでは、1つのサンプルのプローブをアレイエレメントにハイブリダイズし、ハイブリダイゼーション複合体形成後のシグナルを検出する。シグナル強度は、サンプル中のプローブmRNAレベルと相関する。示差ハイブリダイゼーションフォーマットでは、2つの生物サンプルでの遺伝子のセットの示差発現を分析する。2つのサンプルのプローブを調製して、異なる標識成分で標識化する。2つの標識されたプローブの混合液をアレイエレメントにハイブリダイズし、2つの異なる標識からの発光を個々に検出可能な条件下で2つの異なるシグナルを調べる。両方の生物サンプルから由来した実質的に同数のプローブにハイブリダイズするアレイ上のエレメントは、特有の結合した蛍光を出す(Shalon W095/35505)。

10

【0131】

ハイブリダイゼーション複合体は、Cy3の励起のためには488 nm、Cy5の励起のためには632 nmでスペクトル線を発生し得るInnova 70混合ガス10 Wレーザ(Coherent, Santa Clara CA)を備えた顕微鏡で検出する。20 × 顕微鏡対物レンズ(Nikon, Melville NY)を用いて、アレイ上に励起レーザ光の焦点を当てる。アレイを含むスライドを顕微鏡のコンピュータ制御のX-Yステージに置き、20 μ mの解像度で対物レンズを通過してラスタースキャンする。示差ハイブリダイゼーションフォーマットで、レーザにより2つの蛍光色素を同時に励起する。発光された光は、波長に基づき分離され、2つの蛍光色素に対応する2つの光電子増倍管検出器(PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ)に送られる。アレイと光電子増倍管間に設置された好適なフィルターを用いて、シグナルをフィルターする。用いる蛍光色素の最大発光の波長は、Cy3では565 nm、Cy5では650 nmである。プローブ混合液に加えた酵母対照mRNAによって生み出されるシグナル強度を用いてスキャンの感度を校正する。アレイ上の特定の位置には相補的DNA配列を含め、その位置におけるシグナルの強度がハイブリダイズする種の重量比1:100,000に相関するようにする。

20

30

【0132】

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

40

【0133】

7 電子分析

BLASTを用いて、GenBankやLifeSeqデータベース(Incyte Genomics)において同一または関連分子を検索した。ヒト及びラットの配列の積スコアを下記のようにして計算した。BLASTスコアにヌクレオチド一致率を乗じ、積を(2つの配列の短い方の長さの5倍)で除し、短い配列の長さへの100%アラインメントが積スコアを100にする。積スコアは2つの配列間の類似性の程度と配列一致の長さの両方を考慮する。例えば、積スコア40では一致はエラ - が1%から2%以内の厳密さになり、少なくとも積スコア70では一致は厳密となる

50

。類似または関連する分子は、通常、積スコアが8から40の間を示す分子を選択して同定する。

【0134】

電子ノーザン分析を、積スコア70で実施し、図1及び2において示した。LIFESEQデータベースのすべての配列とcDNAライブラリを系、器官/組織、細胞のタイプにより分類した。分類には、心血管系、結合組織、消化器系、胎芽構造、内分泌系、外分泌腺、女性生殖器、男性生殖器、生殖細胞、血液及び免疫系、肝、筋骨格系、神経系、脾臓、呼吸器系、感覚器、皮膚、顎口腔系、非分類性/混合性または尿路がある。各カテゴリーでは、配列が発現したライブラリの数を数えて、そのカテゴリー内のライブラリの総数を示した。非ノーマライズしたライブラリでは、2つ以上の発現レベルが有意である。

10

【0135】

8 相補的分子

cDNAに相補的な分子、約5 (PNA)から5000 bp (cDNAインサートの相補体)を用いて、遺伝子発現を検出あるいは阻害する。検出については、実施例7に記載されている。プロモーターが結合するのを妨げることにより転写を阻害するため、相補的分子を最も固有な5'配列に結合し、オープンリーディングフレームの開始コドンの5' UTR上流のヌクレオチドを含むよう設計する。相補的分子は、ゲノム配列(エンハンサーまたはイントロン等)を含み、ポリメラーゼ、転写因子または調節分子の結合のために十分に開くような二重らせんの能力に影響を及ぼす「三重らせん」塩基対の形成に用いられる。翻訳を阻害するためには、相補的分子を設計して、リボソームが哺乳動物タンパク質をコードするmRNAに結合するのを阻害する。

20

【0136】

相補的分子を発現ベクターに置いて、一過性或いは短期治療向けに効果を試験するために細胞株を器官、腫瘍、滑腔また脈管系に用い、あるいは長期或いは安定した遺伝子治療向けに幹細胞、酵素または他の再生系統に形質転換するために用いる。非複製ベクターで一過性発現は1ヶ月以上続く、また形質転換/発現系でベクター複製を誘導する適切なエレメントが用いられる場合は3ヶ月以上続く。

【0137】

相補的分子をコードするベクターを有する適切な分裂する細胞の安定的な形質転換によって、遺伝形質転換した細胞株、組織または生命体を産生する(USPN 4,736,866)。安定した組込みを可能にする十分な量のベクターを同化、複製する細胞も哺乳動物タンパク質をコードするcDNAの活性に影響を及ぼすか完全に除くための十分な相補的分子を生成する。

30

【0138】

9 配列、マイクロアレイ試薬、及び使用の選択

インサイトクローンは、LIFESEQ GOLDの構築(assembled)されたヒト配列データベース(Incyte Genomics)に由来するテンプレート配列を示すものである。特定のテンプレートに対して1より多いクローンが利用可能であるケースでは、マイクロアレイにテンプレート中の5'-mostクローンが用いられる。HUMAN GENOME GEMシリーズ1-3マイクロアレイ(Incyte Genomics)は、28,626個のアレイエレメントを含み、それは10,068個のannotatedクラスタ及び18,558個のunannotatedクラスタを示す。UNIGEMシリーズのマイクロアレイ(Incyte Genomics)の為には、インサイト社のクローンは非重複性の単一遺伝子クラスタ(Uni gene database (build 46), NCBI; Shuler (1997) J Mol Med 75:694-698)にマッピングされた。また、最も強いBLAST配列(少なくとも90%同一、及び100 bpのオーバーラップ)を有する5'クローンが、選択され、検証され、またマイクロアレイの合成で用いられる。UNIGEM Vマイクロアレイ(Incyte Genomics)は、7075のアレイエレメントを有し、それは4610個のannotated遺伝子及び2,184個のunannotatedクラスタを示す。

40

【0139】

マイクロアレイを構築するためには、cDNAは、cDNA挿入断片をフランキングするベクター配列に対するプライマーの相補性を用いることで細菌細胞から増幅された。30サイクルのPCRは、初期量のcDNA、1~2ngから最終量のcDNA5 µgより多くなるまで増量された。増

50

幅されたcDNAは、次にSEPHACRYL-400カラム (APB) を用いて精製された。精製されたcDNAは、ポリマーコートされたガラススライド上で固定化された。ガラス顕微鏡のスライド (Corning, Corning N.Y.) は、0.1% SDS及びアセトン中で超音波洗浄され、洗浄前後に多量の蒸留水で洗浄された。スライドガラスは、4% フッ化水素酸 (VWR Scientific Products, West Chester PA) 中でエッチングし、蒸留水中で広範囲にわたって洗浄し、95% エタノール中の0.05% アミノプロピルシラン (Sigma Aldrich) でコーティングする。コーティングしたスライドガラスは、110 °Cのオーブンで硬化させる。米国特許第5,807,522号に記載されている方法を用いて、コーティングしたガラス基板にcDNAを付加する。平均濃度が100 ng/μlのcDNA 1 μlを高速機械装置により開放型キャピラリープリンティングエレメント (open capillary printing element) に充填する。次にこの装置が、スライド毎に約5 nlのcDNAを分注する。

10

【0140】

マイクロアレイには、STRATALINKER UVクロスリンカー (Stratagene) を用いてUV架橋する。マイクロアレイは、室温において0.2% SDSで1回洗浄し、蒸留水で3回洗浄する。非特異的な結合部位は、リン酸緩衝生理食塩水 (Tropix, Inc., Bedford MA) における0.2% カゼイン中で60分間マイクロアレイをインキュベートし、その後上述したように0.2% SDS及び蒸留水で洗浄することによってブロックする。

【0141】

1.0 サンプルの調合

LNCaP (ATCC, Manassus VA) は、転移性前立腺癌を患った50歳の男性のリンパ節バイオプシーより単離された前立腺癌株化細胞である。LNCaP細胞は、前立腺特異抗原を発現し、前立腺酸性フォスファターゼを供給し、アンドロゲンレセプタを発現する。LNCaP前立腺癌細胞の遺伝子発現プロファイルは、非腫瘍形成性の原発性 (primary) 前立腺上皮性PrEC細胞と比較された。

20

【0142】

1.1 TRPの発現

哺乳動物細胞発現系か昆虫細胞発現系のどちらかを用いて、タンパク質の発現と精製を達成する。pUB6/V5-His ベクター系 (Invitrogen, Carlsbad CA) を用いて、CHO細胞のCCMを発現する。ベクターは選択可能bsd遺伝子、多数のクローニング部位、ヒトユビキチン C 遺伝子のプロモーター/エンハンサー配列、抗-V5抗体での抗体検出のためのC-末端V5エピトープ、PROBOND 樹脂 (Invitrogen) 上での急速な精製のためのC-末端ポリヒスジン (6xHis) 配列を含む。形質移入した細胞を選択してプラストサイジンを含む培地に移す。

30

【0143】

Spodoptera frugiperda (Sf9) 昆虫細胞を組換え *Autographica californica* 核多角体病ウイルス (AcMNPV) で感染させる。相同組換えによって多角体遺伝子をcDNAと置換する。多角体プロモーターによりcDNAの転写が行われる。上述の精製を可能にする6xhisで、融合タンパク質としてタンパク質を合成する。精製したタンパク質を次の活性で用いて、抗体を作製する。

【0144】

1.2 抗体の産生

ポリアクリルアミドゲル電気泳動法を用いてSTEAPRPを精製して、マウスやウサギを免疫化するために使用する。下記のプロトコルを用いて抗体を産生する。或いは、レーザGENE ソフトウェア (DNASTAR) を用いてSTEAPRPのアミノ酸配列を解析し、免疫原性の高い領域を決定する。通常C-末端付近或いは隣接する親水性領域内で見られる抗原性エピトープを選択して、合成し、抗体を生成するために用いられる。通常は、長さ約15残基のエピトープを、Fmocケミストリーを用いるABI 431A ペプチドシンセサイザ (Applied Biosystems) を用いて生成し、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) に結合させて、免疫原性を高める。

40

【0145】

完全フロイントアジュバントにおいてエピトープ-KLH複合体を用いてウサギを免疫化す

50

る。不完全フロイントアジュバントにおいて免疫化を間隔を置いて反復する。マウスには最低7週間、ウサギには最低12週間後、抗血清を抽出して、抗ペプチド活性のために検査した。検査には、ペプチドをプラスチックに結合すること、1%のウシ血清アルブミンでブロックすること、ウサギ抗血清と反応させて洗浄すること、さらに放射性ヨウ素標識されたヤギ抗ウサギIgGと反応させることが関係する。当分野で公知の方法を用いて、抗体価と形成複合体の量を決定する。

【0146】

1.3 特異的抗体を用いる天然タンパク質の精製

タンパク質を特異結合する抗体を用いて、イムノアフィニティークロマトグラフィにより天然あるいは組換えタンパク質を精製する。イムノアフィニティークラムは、CNBr-活性化SEPHAROSE樹脂 (APB) に抗体を共有結合させることにより形成する。タンパク質を含む培養液をイムノアフィニティークラムに通し、タンパク質を優先的に吸着する条件下 (例えば洗浄剤が存在する高イオン強度緩衝液) でカラムを洗浄する。結合後、そのタンパク質を、抗体とタンパク質との結合を切るために、例えば、pH 2~3のバッファ、或いは高濃度の尿素またはチオシアン酸塩イオンを用いてカラムから溶出させ、タンパク質を回収する。

【0147】

1.4 cDNAまたはタンパク質と特異結合するためのスクリーニング分子

cDNAまたはその断片、タンパク質またはその部分を³²P-dCTP、Cy3-dCTPまたはCy5-dCTP (APB)あるいはBIODIPYかFITC (Molecular Probes, Eugene OR)で標識化する。前もって基質上に配置した候補分子または複合体のライブラリを標識したcDNAまたはタンパク質の存在下でインキュベートする。核酸配列かアミノ酸配列のための条件下でインキュベート後、基質を洗浄し、標識を保持する基質上の、特異結合か複合体成形を示す任意の位置がアッセイされ、リガンドを同定する。異なる濃度の核酸またはタンパク質を用いて得られたデータを使用して、標識された核酸かタンパク質と結合した分子の間の親和性を計算する。

【0148】

1.5 2つのハイブリッドスクリーン

酵母2ハイブリッドシステム、MATCHMAKER LexA 2ハイブリッドシステム (Clontech Laboratories, Palo Alto CA)を使用して、本発明のタンパク質を結合するペプチドをスクリーニングする。タンパク質をコードするcDNAをpLexAベクター、リガンドの多数のクローニング部位に挿入して、大腸菌に形質転換する。mRNAから調製したcDNAをpB42ADベクターの多数のクローニング部位に挿入して、ライゲーションし、cDNAライブラリを作製するために大腸菌に形質転換する。pLexAプラスミドとpB42AD-cDNAライブラリ作製を大腸菌から単離し、ポリエチレングリコール/リチウムアセテートプロトコルを用いて形質転換受容性をもつ酵母EGY48[p8op-lacZ]細胞に同時形質転換するために比率2:1で使用する。形質転換した酵母細胞を、ヒスチジン(-His)、トリプトファン(-Trp)、ウラシル(-Ura)を含まない合成ドロップアウト(SD)培地で培養する、そしてコロニーが成長して数えられるまで、30℃でインキュベートする。コロニーを最小限の容積の1x TE (pH 7.5)で集め、2%ガラクトース(Gal)、1%ラフィノース(Raf)、80 mg/ml 5-ブromo-4-クロロ-3-インドリル-d-ガラクトピラノシド(X-Gal)で補われるSD/-His/-Leu/-Trp/-Ura培地で再培養する。次いで青色コロニーの成長を検査する。発現したタンパク質とcDNA融合タンパク質間の相互作用により、EGY48のLEU2レポーター遺伝子を活性化し、ロイシン(-Leu)を含まない培地でコロニー成長を起こす。相互作用はまた、X-Galで成長するコロニーに青色を産生するp8op-lacZレポーター作成物からの-d-ガラクトシダーゼの発現を活性化する。

【0149】

発現したタンパク質とcDNA融合タンパク質間の陽性の相互作用が、個々の陽性コロニーを単離して、30℃で1から2日間、SD/-Trp/-Ura液体で育成することにより確認できる。培地のサンプルをSD/-Trp/-Ura培地で培養し、コロニーが出現するまで30℃でインキュベートする。サンプルをSD/-Trp/-UraとSD/-His/-Trp/-Uraプレートで複製培養する。ヒスチ

10

20

30

40

50

ジンを含まない培地では成育せず、ヒスチジンを含むSDで成育するコロニーは、pLexAプラスミドを失っている。ヒスチジンを必要とするコロニーをSD/Gal/Raf/X-Gal/-Trp/-Urで成育する。そして白色コロニーを単離して、増殖する。タンパク質と物理的に相互作用するタンパク質をコードするcDNAを含むpB42AD-cDNAプラスミドを酵母細胞から単離して、特徴付ける。

【0150】

1.6 TRPアッセイ

前立腺内に於けるSTEAPRPのローカリゼーションは、Hubertら（既出）による免疫組織化学的な解析によって検出される。前立腺組織切片（4mm）は、ホルマリンで固定され、パラフィン内に包埋された。組織は、抗STEAPRPでインキュベートされ、洗浄され、次にピ

10

【0151】

本明細書において記載した全ての特許出願、刊行物は、言及することをもって本明細書の一部とする。当業者は、本発明の範囲及び精神から逸脱することなく本発明の記載した方法及びシステムの種々の改変を行い得る。本発明について説明するにあたり特定の好適実施例に関連して説明を行ったが、本発明の範囲が、そのような特定の実施例に不当に制限されるべきではないことを理解されたい。実際に、分子生物学または関連分野の専門家には明らかな、本明細書に記載されている本発明の実施方法の様々な改変は、特許請求の範囲内にあるものとする。

20

【0152】

（表の簡単な説明）

表1及び2は、LIFESEQ Goldデータベース（Incyte Genomics, Palo Alto）を用いて作成されたSTEAPRPのノーザン分析を示すものである。表1では第1列が組織カテゴリを示す。第2列は、組織カテゴリ内のトータルクローン数を示す。第3列は、少なくとも一つの転写物が見られるライブラリの数と全体のライブラリ数との比率を示す。第4列は、転写の絶対クローン発生量を示す。第5列は転写の発生量のパーセンテージを示す。表2は、とりわけ癌患者の前立腺組織内のSTEAPRP発現を示すものである。第1列は、ライブラリ名をリストしている。第2列はライブラリのために配列されたクローンの数を示している。第3列はライブラリが誘導された組織について記載している。第4列は転写の絶対的な発生量を示す。第5列は、転写の発生量のパーセンテージを示す。

30

【0153】

表3は、マイクロアレイ解析によるヒトPrEC非腫瘍形成性前立腺上皮性細胞と比較したヒトLNCaP前立腺癌細胞におけるSTEAPRPの差次的発現を示す。第1列はlog2 DE（LNCaP細胞/PrEC細胞）として示された平均の差次的発現（DE）値をリストしている。第2列は、差次的発現値における共分散パーセンテージをリストしている。第4列は、蛍光レッドダイCy5でラベルされたLNCaP-誘導サンプルをリストしている。

【図面の簡単な説明】

【0154】

【図1-A】図は、cDNA（SEQ ID NO:2）によってコードされたSTEAPRP（SEQ ID NO:1）を示している。翻訳は、MACDNASIS PRO software（Hitachi Software Engineering, South San Francisco CA.）によってなされた。

40

【図1-B】図は、cDNA（SEQ ID NO:2）によってコードされたSTEAPRP（SEQ ID NO:1）を示している。翻訳は、MACDNASIS PRO software（Hitachi Software Engineering, South San Francisco CA.）によってなされた。

【図1-C】図は、cDNA（SEQ ID NO:2）によってコードされたSTEAPRP（SEQ ID NO:1）を示している。翻訳は、MACDNASIS PRO software（Hitachi Software Engineering, South San Francisco CA.）によってなされた。

【図1-D】図は、cDNA（SEQ ID NO:2）によってコードされたSTEAPRP（SEQ ID NO:1）を示している。翻訳は、MACDNASIS PRO software（Hitachi Software Engineering, South Sa

50

n Francisco CA.)によってなされた。

【図 1 - E】図は、cDNA(SEQ ID NO:2)によってコードされたSTEAPRP(SEQ ID NO: 1)を示している。翻訳は、MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA.)によってなされた。

【図 1 - F】図は、cDNA(SEQ ID NO:2)によってコードされたSTEAPRP(SEQ ID NO: 1)を示している。翻訳は、MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA.)によってなされた。

【図 1 - G】図は、cDNA(SEQ ID NO:2)によってコードされたSTEAPRP(SEQ ID NO: 1)を示している。翻訳は、MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA.)によってなされた。

10

【図 2 - A】図は、STEAPRP(7492448; SEQ ID NO: 1)及びヒトSTEAP (g6572948; SEQ ID NO:11)の配列及びドメイン間の保存された化学的、構造的類似性を示している。配列は、LASERGENEソフトウェアのMEGALIGNプログラム(DNASTAR, Madison WI)によってなされた。

【図 2 - B】図は、STEAPRP(7492448; SEQ ID NO: 1)及びヒトSTEAP (g6572948; SEQ ID NO:11)の配列及びドメイン間の保存された化学的、構造的類似性を示している。配列は、LASERGENEソフトウェアのMEGALIGNプログラム(DNASTAR, Madison WI)によってなされた。

【図 2 - C】図は、STEAPRP(7492448; SEQ ID NO: 1)及びヒトSTEAP (g6572948; SEQ ID NO:11)の配列及びドメイン間の保存された化学的、構造的類似性を示している。配列は、LASERGENEソフトウェアのMEGALIGNプログラム(DNASTAR, Madison WI)によってなされた。

【 0 1 5 5 】

20

【表 1】

表1

組織カテゴリ	クローン カウント	発見された数	絶対 発生量	発生量の パーセンテージ
心血管系	266190	0/68	0	0.0000
結合組織	144645	0/47	0	0.0000
消化器系	501101	3/148	3	0.0006
胚構造	106713	0/21	0	0.0000
内分泌系	225386	2/53	2	0.0009
外分泌腺	254635	1/64	1	0.0004
生殖 女性	427284	2/106	2	0.0005
生殖 男性	448207	28/114	43	0.0096
生殖細胞	38282	0/5	0	0.0000
血液及び免疫系	680277	2/159	3	0.0004
肝臓	109378	1/35	2	0.0018
筋骨格系	159280	2/47	3	0.0019
神経系	955753	9/198	12	0.0013
脾臓	110207	1/24	2	0.0018
呼吸器系	390086	6/93	9	0.0023
感覚器	19256	0/8	0	0.0000
皮膚	72292	0/15	0	0.0000
顎口腔系	12923	0/10	0	0.0000
非分類/混合	120926	1/13	1	0.0008
尿路	279062	2/64	2	0.0007
合計	5321883	60/1292	85	0.0000

30

【 0 1 5 6 】

40

【表 2】

10

20

30

【 0 1 5 7 】
【 表 3 】

表2

ライブラリID	クローン カウント	ライブラリの解説	絶対 発生量	発生量の ハーゼンテージ
PROSNOT19	3679	prostate, AH, mw/adenocA, M	4	0.1087
PROSTUT18	2201	prostate tumor, adenocA, 68M	2	0.0909
PROETWP06	1157	prostate, PIN, mw/cancer, M	1	0.0864
PROSNOT26	3705	prostate, mw/adenocA, 65M	3	0.0810
PROSDIT01	3873	prostate, AH, mw/adenocA, 58M	3	0.0775
PROSTUS20	4550	prostate tumor, adenocA, 59M, SUB	3	0.0659
PROSTUT04	8553	prostate tumor, adenocA, 57M	3	0.0351
PROSNOT20	2995	prostate, AH, mw/adenocA, 65M	1	0.0334
PROSTUT21	3268	prostate tumor, adenocA, 61M	1	0.0306
PROSTUS19	4087	prostate tumor, adenocA, 59M, SUB	1	0.0245
PROSTUT12	7138	prostate tumor, adenocA, 65M	1	0.0140
PROSNOT06	8829	prostate, AH, mw/adenocA, 57M	1	0.0113

表3

mean log2 DE (Cy5/Cy3)	CV %	Cy3	Cy5
1.82	0	Human, P1EC Cells, Starved 24hr	Human, LNCaP Line, Starved 24hr, CA
3	0	Human, P1EC Cells, Starved 24hr	Human, LNCaP Line, Starved 24hr, CA
1.86	0	Human, P1EC Cells, Starved 24hr	Human, LNCaP Line, Rich media 24hr, CA
3.02	0	Human, P1EC Cells, Starved 24hr	Human, LNCaP Line, Rich media 24hr, CA
2.5	12.45	Human, P1EC Cells	Human, LNCaP Line, CA

【 図 1 - A 】

```

5'  9   18   27   36   45   54
    GG GGA AGC AGC TGG AGT GCG ACC AGC GCA ACC ACC CTG CAA CCG CCA GTC
    63   72   81   90   99   108
    GGA GGT GCA GTC CGT AGG CCC CTG CCC GGT GGG CCC TTG GGG AGT CCG CCG
    117  126  135  144  153  162
    CCG TCC CGA GGA GCT GCA AGG CTC GCC CCG CCG CCG GGT GGA GGG CCG GGG
    171  180  189  198  207  216
    CCG GGA GAA AGT GAA GAG AGG AAA TTG GAA AAT TGT GAG TGG ACC TTC TGA TAC
    225  234  243  252  261  270
    TGC TCC TCC TTG CCG GGA AAA GGG GAA AGA ACT GCA TGC ATA TTA TTC AGC GTC
    279  288  297  306  315  324
    CTA TAT TCA AAG GAT ATT CTT GGT GAT CTT GGA AGT GTC CGT ATC AAG GAA TCA
    333  342  351  360  369  378
    ATC TCT ATG ATG GGA AGC CCT AAG AGC CTT AGT GAA ACT TGT TTA CCT AAT GGC
    I S M M G S P K S L S E T C L P N G

```

【 図 1 - B 】

```

387  396  405  414  423  432
    ATA AAT GGT ATC AAA GAT GCA AGG AAG GTC ACT GTA GGT GTG ATT GGA AGT GGA
    I N G I K D A R K V T V G V I G S G
    441  450  459  468  477  486
    GAT TTT GCC AAA TCC TTG ACC ATT CGA CTT ATT AGA TGC GGC TAT CAT GTG GTC
    D F A K S L T I R L I R C G Y H V
    495  504  513  522  531  540
    ATA GGA AGT AGA AAT CCT AAG TTT GCT TCT GAA TTT TTT CCT CAT GTG GTA GAT
    I G S R N P K F A S E F F P H V D
    549  558  567  576  585  594
    GTC ACT CAT CAT GAA GAT GCT CTC ACA AAA ACA AAT ATA ATA TTT GTT GCT ATA
    V T H H E D A L T K T N I I F V A I
    603  612  621  630  639  648
    CAC AGA GAA CAT TAT ACC TCC CTG TGG GAC CTG AGA CAT CTG CTT GTG GGT AAA
    H R E H Y T S L W D L R H L L V G R

```

【 図 1 - C 】

```

657  666  675  684  693  702
    ATC CTG ATT GAT GTG AGC AAT AAC ATG AGG ATA AAC CAG TAC CCA GAA TCC AAT
    I L I D V S N N M R I N Q Y P E S N
    711  720  729  738  747  756
    GCT GAA TAT TTG GCT TCA TTA TTC CCA GAT TCT TTG ATT GTC AAA GGA TTT AAT
    A E Y L A S L F D S L I V K G F N
    765  774  783  792  801  810
    GTT GTC TCA GCT TGG GCA CTT CAG TTA GGA CCT AAG GAT GCC AGC CAG GTT
    V V S A W A L Q L G P K D A S R Q
    819  828  837  846  855  864
    TAT ATA TGC AGC AAC AAT ATT CAA GCG CGA CAA CAG GTT ATT GAA CTT GCC CCG
    Y I C S N N I Q A R Q Q V I E L A R
    873  882  891  900  909  918
    CAG TTG AAT TTC ATT CCC ATT GAC TTG GGA TCC TTA TCA TCA GCC AGA GAG ATT
    Q L N F I P I D L G S L S S A R E I

```

【 図 1 - D 】

```

927  936  945  954  963  972
    GAA AAT TTA CCC CTA CGA CTC TTT ACT CTC TCG AGA GGG CCA GAT GTG GTA GCT
    E N L P L R L F T L W R G P V V A
    981  990  999  1008  1017  1026
    ATA AGC TTG GCC ACA TTT TTT TTC CTT TAT TCC TTT GTC AGA GAT GTG AAT CAT
    I S L A T F F L Y S F V R D V I H
    1035  1044  1053  1062  1071  1080
    CCA TAT GCT AGA AAC CAA CAG AGT GAC TTT TAC AAA ATT CCT ATA GAG ATT GTG
    P Y A R N Q Q S D F Y K I P I E I V
    1089  1098  1107  1116  1125  1134
    AAT AAA ACC TTA CCT ATA GTT GCC ATT ATG CTC TCC CTA GTA TAC CTC GCA
    N K T L P I V A I T L L S L V Y L A
    1143  1152  1161  1170  1179  1188
    GGT CTT CTG GCA GCT GCT TAT CAA CTT TAT TAC GGC ACC AAG TAT AGG AGA TTT
    G L L A A A Y Q L Y Y G T K Y R R F

```

【 図 1 - E 】

```
1197 1206 1215 1224 1233 1242
CCA CCT TGG TTG GAA ACC TGG TTA CAG TGT AGA AAA CAG CTT GGA TTA CTA AGT
P P W L E T W L Q C R K Q L G L L S
1251 1260 1269 1278 1287 1296
TTT TTC TTT GCT ATG GTC CAT GTT GCC TAC AGC CTC TGC TTA CCG ATG AGA AGG
F F A M V H V A Y S L C L P M R R
1305 1314 1323 1332 1341 1350
TCA GAG AGA TAT TTG TTT CTC AAC ATG GCT TAT CAG CAG GTT CAT GCA AAT ATT
S E R Y L F L N M A Y Q Q V H A N I
1359 1368 1377 1386 1395 1404
GAA AAC TCT TGG AAT GAG GAA GAA GTT TGG AGA ATT GAA ATG TAT ATC TCC TTT
E N S W N E E V W R I E M Y I S F
1413 1422 1431 1440 1449 1458
GGC ATA ATG AGC CTT GGC TTA CTT TGC CTC CTG GCA GTC ACT TCT ATC CCT TCA
G I M S L G L L S L L A V T S I P S
```

【 図 1 - F 】

```
1467 1476 1485 1494 1503 1512
GTG AGC AAT GCT TTA AAC TGG AGA GAA TTC AGT TTT ATT CAG TCT ACA CTT GGA
V S N A L N W R E F S F I Q S T L G
1521 1530 1539 1548 1557 1566
TAT GTC GCT CTG CTC ATA AGT ACT TTC CAT GTT TTA ATT TAT GGA TGG AAA CGA
Y V A L L I S T F H V L I Y G W K R
1575 1584 1593 1602 1611 1620
GCT TTT GAG GAA GAG TAC TAC AGA TTT TAT ACA CCA AAC TAT TTT GTT CTT GCT
A F E E E Y Y R F Y T P P N F V L A
1629 1638 1647 1656 1665 1674
CTT GTT TTG CCC TCA ATT GTA ATT CTG GGT AAG ATT ATT TTA TTC CTT CCA TGT
L V L P S I V I L G K I I L F L P C
1683 1692 1701 1710 1719 1728
ATA AGC CGA AAG CTA AAA CGA ATT AAA AAA GGC TGG GAA AAG AGC CAA TTT CTG
I S R K L K R I K K G W E K S Q F L
```

【 図 1 - G 】

```
1737 1746 1755 1764 1773 1782
GAA GAA GGT ATT GGA GGA ACA ATT CCT CAT GTC TCC CCG GAG AGG GTC ACA GTA
E E G I G T I P H V S P E R V T V
1791 1800 1809 1818 1827 1836
ATG TGA TGA TAA ATG GTC TTC ACA GCT GCC ATA TAA AGT TCT ACT CAT GCC ATT
M
1845 1854 1863 1872 1881 1890
ATT TTT ATG ACT TCT ACG TTC AGT TAC AAG TAT GCT GTC AAA TTA TCG TGG GTT
GA 3'
---
```

【 図 2 - A 】

```
1 MES I S M M G S P K S L S E T C L P N G I N G I K D A R K 7492448
1 MES
31 VT V G V I G S G D F A K S L T I R L I R C G Y H V V I G S 7492448
6
61 R N P K F A S E F F P H V V D V T H H E D A L T K T N I I F 7492448
6
91 V A I H R E H Y T S L W D L R H L L V G K I L I D V S N N M 7492448
13
121 R I N Q Y P E S N A E Y L A S L F P D S L I V K G F N V S 7492448
23 E E D D Y L H K D T G E T S M L K R P V L
151 A W A L Q L G P K D A S R Q V Y I C I S N N I Q A R Q V I E 7492448
44 - - - L H L - H Q T A H A D E F D C P S E L Q H T Q E - - - 7492448
181 L A R Q L N F I P I D L G S L S A R E I E N L P L R L F T 7492448
67 - - - - - - - - - - - - - - - L F P 7492448
211 L W R G P V V V A I S L A T F F F L Y S F V R D V I H P Y A 7492448
70 Q W H L P I K I A I A S L T F L Y T L L R E V I H P L A 7492448
```

【 2 - B 】

241 RNQSDFYKIPITIEIVNK(TLP)I(VA)ITLLS(LV) 7492448
100 TSHQOQYFYKIPITILVINKVLLPMVSI(TLLA)LV 96572948
271 YLA(G)LLAAY(Q)LYYG(TKY)RR(FP)W(L)E(TW)LO 7492448
130 YLP(GV)IAAIVQLHNGTKYKKFP(HWL)DK(WML) 96572948
301 CRKQ(L)GLLSFFFA(MV)H(VA)YS(L)CL(PMRR)S(E)R 7492448
160 TRKQFGLLSFFFAVLHAIYSLSYPMRRSYR 96572948
331 YLFLNMAY(Q)V(HA)N(I)ENS(WN)EE(VWR)I(E)MY 7492448
190 YKLLNWAY(Q)VQNKEDAWIEHDVWRMEIY 96572948
361 ISFGIMSL(G)L(LS)LLAVTSIPSVSNA(L)NWRE 7492448
220 VSLGIVGLA(LA)LLAVTSIPSVSDSLTWRE 96572948
391 F(SFI)QSTLGY(VA)LLIS(T)E(HV)LYG(WKRA)FE 7492448
250 FHYIQSKLGIYSLLLGTIHALIFAWNKKI(D) 96572948
421 EYYR(FY)TP(NFV)LA(LVLP)SIV(ILGK)I(LF) 7492448
280 IKQFVWY(TP)PTFMIAVFLPIVVLIFKSI(LF) 96572948

【 2 - C 】

451 L P C I S R R K L K R I K K G W E K S Q F L E E G I G G T I P 7492448
310 L P C L R K K I L K I R H G W E D - - - - - V T 96572948
481 H V S P E R V - T V M 7492448
329 K I N K T E I C S Q L 96572948

【国際公開パンフレット】

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

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

PCT

(10) International Publication Number
WO 02/072596 A1(51) International Patent Classification: C07H 21/02,
21/04, C12P 21/06, C12N 5/00

(21) International Application Number: PCT/US02/07053

(22) International Filing Date: 7 March 2002 (07.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data: 09/802,520 9 March 2001 (09.03.2001) US

(71) Applicant (for all designated States except US): INCYTE
GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo
Alto, CA 94304 (US).(72) Inventors; and
(75) Inventors/Applicants (for US only): LAL, Preeti, G.
[IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US).
FARIS, Mary [US/US]; 2538 Almaden Court, Los An-
geles, CA 90077 (US). CHEN, Hwei-Mei [US/US]; 2425
Lyle Court, Santa Clara, CA 95056 (US). ISON, Craig,
H. [US/US]; 1242 Wenthersfield Way, San Jose, CA 95118
(US).(74) Agents: STREETER, David, G. et al.; Incyte Genomics,
Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).(81) Designated States (national): AE, AG, AI, 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, LI, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MY, NZ, OM, PH, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VN, YU, ZA, ZM, ZW.(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,
GB, GR, IE, IT, LU, MC, NL, PT, SI, TR), OAPI patent
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,
NE, SN, TD, TG).**Published:**
with international search report
— before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendmentsFor two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

WO 02/072596 A1

(54) Title: STEAP-RELATED PROTEIN

(57) Abstract: The invention provides a cDNA which encodes a STIAP-related protein. It also provides for the use of the cDNA, fragments, complements, and variants thereof and of the encoded protein, portions thereof and antibodies thereto for diagnosis and treatment of prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer. The invention additionally provides expression vectors and host cells for the production of the protein and a transgenic model system.

WO 02/072596

PCT/US02/07053

STEAP-RELATED PROTEIN

TECHNICAL FIELD

5 This invention relates to cDNA which encodes a STEAP-related protein and to the use of the cDNA and the encoded protein in the diagnosis and treatment of prostate cell proliferative disorders, in particular, prostate hyperplasia and prostate cancer.

BACKGROUND OF THE INVENTION

Phylogenetic relationships among organisms have been demonstrated many times, and studies
10 from a diversity of prokaryotic and eukaryotic organisms suggest a more or less gradual evolution of molecules, biochemical and physiological mechanisms, and metabolic pathways. Despite different evolutionary pressures, the proteins of nematode, fly, rat, and man have common chemical and structural features and generally perform the same cellular function. Comparisons of the nucleic acid and protein sequences from organisms where structure and/or function are known accelerate the
15 investigation of human sequences and allow the development of model systems for testing diagnostic and therapeutic agents for human conditions, diseases, and disorders.

Prostate cancer is a common malignancy in men over the age of 50, and the incidence increases with age. In the US, there are approximately 132,000 newly diagnosed cases of prostate cancer and more than 33,000 deaths from the disorder each year. Once cancer cells arise in the
20 prostate, they are stimulated by testosterone to a more rapid growth. Thus, removal of the testes can indirectly reduce both rapid growth and metastasis of the cancer. Over 95 percent of prostatic cancers are adenocarcinomas which originate in the prostatic acini. The remaining 5 percent are divided between squamous cell and transitional cell carcinomas, both of which arise in the prostatic ducts or other parts of the prostate gland.

25 As with most cancers, prostate cancer develops through a multistage progression ultimately resulting in an aggressive, metastatic phenotype. The initial step in tumor progression involves the hyperproliferation of normal luminal and/or basal epithelial cells that become hyperplastic and evolve into early-stage tumors. The early-stage tumors are localized in the prostate but eventually may metastasize, particularly to the bone, brain, or lung. About 80% of these tumors remain responsive to
30 androgen treatment, an important hormone controlling the growth of prostate epithelial cells. However, in its most advanced state, cancer growth becomes androgen-independent and there is currently no known treatment for this condition.

A primary diagnostic marker for prostate cancer is prostate specific antigen (PSA). PSA is a tissue-specific serine protease almost exclusively produced by prostatic epithelial cells. The quantity of PSA correlates with the number and volume of the prostatic epithelial cells, and consequently, the levels of PSA are an excellent indicator of abnormal prostate growth. Men with prostate cancer

WO 02/072596

PCT/US02/07053

exhibit an early linear increase in PSA levels followed by an exponential increase prior to diagnosis. However, since PSA levels are also influenced by factors such as inflammation, androgen and other growth factors, some scientists and clinicians maintain that changes in PSA levels are not useful in detecting individual cases of prostate cancer.

5 Current areas of cancer research provide additional prospects for markers as well as potential therapeutic targets for prostate cancer. Several growth factors have been shown to play a critical role in tumor development, growth, and progression. The growth factors epidermal growth factor (EGF), fibroblast growth factor (FGF), and transforming growth factor alpha (TGF α) are important in the growth of normal as well as hyperproliferative prostate epithelial cells, particularly at early stages of
10 tumor development and progression, and affect signaling pathways in these cells in various ways (Lin *et al.* (1999) *Cancer Res* 59:2891-2897; Putz *et al.* (1999) *Cancer Res* 59:227-233). The TGF- β family of growth factors are generally expressed at increased levels in human cancers and the high expression levels in many cases correlates with advanced stages of malignancy and poor survival (Gold (1999) *Crit Rev Oncog* 10:303-360). Finally, there are human cell lines representing both the
15 androgen-dependent stage of prostate cancer (LNCap) as well as the androgen-independent, hormone refractory stage of the disease (PC3 and DU-145) that have proved useful in studying gene expression patterns associated with the progression of prostate cancer, and the effects of cell treatments on these expressed genes (Chung (1999) *Prostate* 38:199-207).

Six-transmembrane epithelial antigen of the prostate (STEAP) is a prostate-specific cell-
20 surface marker (Hubert *et al.* (1999) *Proc Natl Acad Sci* 96:14523-14528). STEAP is 339 amino acids in length and has six predicted membrane-spanning regions. It is highly expressed in normal and cancerous prostate tissues and in several prostate cancer-derived cell lines. Its level of expression is insensitive to the presence of androgen. Immunostaining shows that STEAP is located at the plasma membrane of prostate cells where it concentrates at cell-cell junctions of the secretory epithelium.
25 Cell surface antigens such as STEAP may be useful in antibody therapy, cancer-vaccines, and diagnostic imaging for treatment of prostate cancer.

The discovery of a cDNA encoding STEAP-related protein satisfies a need in the art by providing compositions which are useful in the diagnosis and treatment of prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer.

30 SUMMARY OF THE INVENTION

The invention is based on the discovery of a cDNA encoding STEAP-related protein (STEAPRP) which is useful in the diagnosis and treatment of prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer.

The invention provides an isolated cDNA comprising a nucleic acid sequence encoding a
35 protein having the amino acid sequence of SEQ ID NO:1. The invention also provides an isolated

WO 02/072596

PCT/US02/07053

cDNA or the complement thereof selected from the group consisting of a nucleic acid sequence of SEQ ID NO:2, a fragment of SEQ ID NO:2 selected from SEQ ID NOs:3-9, and a variant of SEQ ID NO:2, SEQ ID NO:10. The invention additionally provides a composition, a substrate, and a probe comprising the cDNA, or the complement of the cDNA, encoding STEAPRP. The invention further provides a vector containing the cDNA, a host cell containing the vector and a method for using the cDNA to make STEAPRP. The invention still further provides a transgenic cell line or organism comprising the vector containing the cDNA encoding STEAPRP. The invention additionally provides a fragment, a variant, or the complement of the cDNA selected from the group consisting of SEQ ID NOs:2-10. In one aspect, the invention provides a substrate containing at least one of these fragments or variants or the complements thereof. In a second aspect, the invention provides a probe comprising a cDNA or the complement thereof which can be used in methods of detection, screening, and purification. In a further aspect, the probe is a single-stranded complementary RNA or DNA molecule.

The invention provides a method for using a cDNA to detect the differential expression of a nucleic acid in a sample comprising hybridizing a probe to the nucleic acids, thereby forming hybridization complexes and comparing hybridization complex formation with a standard, wherein the comparison indicates the differential expression of the cDNA in the sample. In one aspect, the method of detection further comprises amplifying the nucleic acids of the sample prior to hybridization. In another aspect, the method showing differential expression of the cDNA is used to diagnose prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer. In another aspect, the cDNA or a fragment or a variant or the complements thereof may comprise an element on an array.

The invention additionally provides a method for using a cDNA or a fragment or a variant or the complements thereof to screen a library or plurality of molecules or compounds to identify at least one ligand which specifically binds the cDNA, the method comprising combining the cDNA with the molecules or compounds under conditions allowing specific binding, and detecting specific binding to the cDNA, thereby identifying a ligand which specifically binds the cDNA. In one aspect, the molecules or compounds are selected from aptamers, DNA molecules, RNA molecules, peptide nucleic acids, artificial chromosome constructions, peptides, transcription factors, repressors, and regulatory molecules.

The invention provides a purified protein or a portion thereof selected from the group consisting of an amino acid sequence of SEQ ID NO:1, a variant having at least 55% identity to the amino acid sequence of SEQ ID NO:1, an antigenic epitope of SEQ ID NO:1, and a biologically active portion of SEQ ID NO:1. The invention also provides a composition comprising the purified protein in conjunction with a pharmaceutical carrier. The invention further provides a method of using the

WO 02/072596

PCT/US02/07053

STEAPRP to treat a subject with prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer comprising administering to a patient in need of such treatment the composition containing the purified protein. The invention still further provides a method for using a protein to screen a library or a plurality of molecules or compounds to identify at least one ligand, the method
5 comprising combining the protein with the molecules or compounds under conditions to allow specific binding and detecting specific binding, thereby identifying a ligand which specifically binds the protein. In one aspect, the molecules or compounds are selected from DNA molecules, RNA molecules, peptide nucleic acids, peptides, proteins, mimetics, agonists, antagonists, antibodies, immunoglobulins, inhibitors, and drugs. In another aspect, the ligand is used to treat a subject with
10 prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer.

The invention provides a method of using a protein to screen a subject sample for antibodies which specifically bind the protein comprising isolating antibodies from the subject sample, contacting the isolated antibodies with the protein under conditions that allow specific binding, dissociating the antibody from the bound-protein, and comparing the quantity of antibody with known
15 standards, wherein the presence or quantity of antibody is diagnostic of prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer.

The invention also provides a method of using a protein to prepare and purify antibodies comprising immunizing a animal with the protein under conditions to elicit an antibody response, isolating animal antibodies, attaching the protein to a substrate, contacting the substrate with isolated
20 antibodies under conditions to allow specific binding to the protein, dissociating the antibodies from the protein, thereby obtaining purified antibodies.

The invention provides a purified antibody which binds specifically to a protein which is expressed in prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer. The invention also provides a method of using an antibody to diagnose prostate cell proliferative
25 disorders, particularly prostate hyperplasia and prostate cancer comprising combining the antibody comparing the quantity of bound antibody to known standards, thereby establishing the presence of prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer. The invention further provides a method of using an antibody to treat prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer comprising administering to a patient in need of
30 such treatment a pharmaceutical composition comprising the purified antibody.

The invention provides a method for inserting a heterologous marker gene into the genomic DNA of a mammal to disrupt the expression of the endogenous polynucleotide. The invention also provides a method for using a cDNA to produce a mammalian model system, the method comprising constructing a vector containing the cDNA selected from SEQ ID NOs:2-10, transforming the vector
35 into an embryonic stem cell, selecting a transformed embryonic stem, microinjecting the transformed

WO 02/072596

PCT/US02/07053

embryonic stem cell into a mammalian blastocyst, thereby forming a chimeric blastocyst, transferring the chimeric blastocyst into a pseudopregnant dam, wherein the dam gives birth to a chimeric offspring containing the cDNA in its germ line, and breeding the chimeric mammal to produce a homozygous, mammalian model system.

5 BRIEF DESCRIPTION OF THE FIGURES AND TABLES

Figures 1A, 1B, 1C, 1D, 1E, 1F, and 1G show the STEAPRP (SEQ ID NO:1) encoded by the cDNA (SEQ ID NO:2). The translation was produced using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA).

Figures 2A, 2B, and 2C demonstrate the conserved chemical and structural similarities among 10 the sequences and domains of STEAPRP (7492448; SEQ ID NO:1) and human STEAP (g6572948; SEQ ID NO:11). The alignment was produced using the MEGALIGN program of LASERGENE software (DNASTAR, Madison WI).

Tables 1 and 2 show the northern analysis for STEAPRP produced using the LIFESEQ Gold database (Incyte Genomics, Palo Alto CA). In Table 1, the first column presents the tissue categories; 15 the second column, the total number of clones in the tissue category; the third column, the ratio of the number of libraries in which at least one transcript was found to the total number of libraries; the fourth column, absolute clone abundance of the transcript; and the fifth column, percent abundance of the transcript. Table 2 shows expression of STEAPRP in prostate tissues, particularly from patients with cancer. The first column lists the library name, the second column, the number of clones 20 sequenced for that library; the third column, the description of the tissue from which the library was derived; the fourth column, the absolute abundance of the transcript; and the fifth column, the percent abundance of the transcript.

Table 3 shows the differential expression of STEAPRP in human LNCaP prostate carcinoma cells compared to human PrEC nontumorigenic prostate epithelial cells as determined by microarray 25 analysis. Column 1 lists the mean differential expression (DE) values presented as log₂ DE (LNCaP cells/PrEC cells). Column 2 lists the percentage covariance (CV%) in differential expression values. Column 3 lists the PrEC-derived samples labeled with fluorescent green dye Cy3. Column 4 lists the LNCaP-derived samples labeled with fluorescent red dye Cy5.

DESCRIPTION OF THE INVENTION

30 It is understood that this invention is not limited to the particular machines, materials and methods described. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments and is not intended to limit the scope of the present invention which will be limited only by the appended claims. As used herein, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. For example, a reference 35 to "a host cell" includes a plurality of such host cells known to those skilled in the art.

WO 02/072596

PCT/US02/07053

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. 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

"STEAPRP" refers to a purified protein obtained from any mammalian species, including bovine, canine, murine, ovine, porcine, rodent, simian, and preferably the human species, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

"Array" refers to an ordered arrangement of at least two cDNAs on a substrate. At least one of the cDNAs represents a control or standard, and the other, a cDNA of diagnostic or therapeutic interest. The arrangement of from about two to about 40,000 cDNAs on the substrate assures that the size and signal intensity of each labeled hybridization complex formed between each cDNA and at least one sample nucleic acid is individually distinguishable.

The "complement" of a cDNA of the Sequence Listing refers to a nucleic acid molecule which is completely complementary over its full length and which will hybridize to the cDNA or an mRNA under conditions of maximal stringency.

"cDNA" refers to an isolated polynucleotide, nucleic acid molecule, or any fragment or complement thereof. It may have originated recombinantly or synthetically, may be double-stranded or single-stranded, represents coding and noncoding 3' or 5' sequence, and generally lacks introns.

The phrase "cDNA encoding a protein" refers to a nucleotide sequence that closely aligns with sequences which encode conserved regions, motifs or domains that were identified by employing analyses well known in the art. These analyses include BLAST (Basic Local Alignment Search Tool) which provides identity within the conserved region (Altschul (1993) J Mol Evol 36: 290-300; Altschul et al. (1990) J Mol Biol 215:403-410).

A "composition" comprises the polynucleotide and a labeling moiety or a purified protein in conjunction with a pharmaceutical carrier.

"Derivative" refers to a cDNA or a protein that has been subjected to a chemical modification. Derivatization of a cDNA can involve substitution of a nontraditional base such as queosine or of an analog such as hypoxanthine. These substitutions are well known in the art. Derivatization of a protein involves the replacement of a hydrogen by an acetyl, acyl, alkyl, amino, formyl, or morpholino group. Derivative molecules retain the biological activities of the naturally occurring molecules but may confer advantages such as longer lifespan or enhanced activity.

"Differential expression" refers to an increased, upregulated or present, or decreased,

WO 02/072596

PCT/US02/07053

downregulated or absent, gene expression as detected by presence, absence or at least two-fold changes in the amount of transcribed messenger RNA or translated protein in a sample.

"Disorder" refers to conditions, diseases or syndromes in which the cDNAs and STEAPRP are differentially expressed. Such a disorder includes prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer.

"Fragment" refers to a chain of consecutive nucleotides from about 50 to about 4000 base pairs in length. Fragments may be used in PCR or hybridization technologies to identify related nucleic acid molecules and in binding assays to screen for a ligand. Such ligands are useful as therapeutics to regulate replication, transcription or translation.

A "hybridization complex" is formed between a cDNA and a nucleic acid of a sample when the purines of one molecule hydrogen bond with the pyrimidines of the complementary molecule, e.g., 5'-A-G-T-C-3' base pairs with 3'-T-C-A-G-5'. Hybridization conditions, degree of complementarity and the use of nucleotide analogs affect the efficiency and stringency of hybridization reactions.

"Labeling moiety" refers to any visible or radioactive label than can be attached to or incorporated into a cDNA or protein. Visible labels include but are not limited to anthocyanins, green fluorescent protein (GFP), β glucuronidase, luciferase, Cy3 and Cy5, and the like. Radioactive markers include radioactive forms of hydrogen, iodine, phosphorous, sulfur, and the like.

"Ligand" refers to any agent, molecule, or compound which will bind specifically to a polynucleotide or to an epitope of a protein. Such ligands stabilize or modulate the activity of polynucleotides or proteins and may be composed of inorganic and/or organic substances including minerals, cofactors, nucleic acids, proteins, carbohydrates, fats, and lipids.

"Oligonucleotide" refers a single-stranded molecule from about 18 to about 60 nucleotides in length which may be used in hybridization or amplification technologies or in regulation of replication, transcription or translation. Substantially equivalent terms are amplimer, primer, and oligomer.

"Portion" refers to any part of a protein used for any purpose; but especially, to an epitope for the screening of ligands or for the production of antibodies.

"Post-translational modification" of a protein can involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and the like. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cellular location, cell type, pH, enzymatic milieu, and the like.

"Probe" refers to a cDNA that hybridizes to at least one nucleic acid in a sample. Where targets are single-stranded, probes are complementary single strands. Probes can be labeled with reporter molecules for use in hybridization reactions including Southern, northern, *in situ*, dot blot, array, and like technologies or in screening assays.

WO 02/072596

PCT/US02/07053

"Protein" refers to a polypeptide or any portion thereof. A "portion" of a protein refers to that length of amino acid sequence which would retain at least one biological activity, a domain identified by PFAM or PRINTS analysis or an antigenic epitope of the protein identified using Kyte-Doolittle algorithms of the PROTEAN program (DNASTAR, Madison WI). An "oligopeptide" is an amino acid sequence from about five residues to about 15 residues that is used as part of a fusion protein to produce an antibody.

"Purified" refers to any molecule or compound that is separated from its natural environment and is from about 60% free to about 90% free from other components with which it is naturally associated.

"Sample" is used in its broadest sense as containing nucleic acids, proteins, antibodies, and the like. A sample may comprise a bodily fluid; the soluble fraction of a cell preparation, or an aliquot of media in which cells were grown; a chromosome, an organelle, or membrane isolated or extracted from a cell; genomic DNA, RNA, or cDNA in solution or bound to a substrate; a cell; a tissue; a tissue print; a fingerprint, buccal cells, skin, or hair; and the like.

"Specific binding" refers to a special and precise interaction between two molecules which is dependent upon their structure, particularly their molecular side groups. For example, the intercalation of a regulatory protein into the major groove of a DNA molecule or the binding between an epitope of a protein and an agonist, antagonist, or antibody.

"Similarity" as applied to sequences, refers to the quantification (usually percentage) of nucleotide or residue matches between at least two sequences aligned using a standardized algorithm such as Smith-Waterman alignment (Smith and Waterman (1981) *J Mol Biol* 147:195-197) or BLAST2 (Altschul *et al.* (1997) *Nucleic Acids Res* 25:3389-3402). BLAST2 may be used in a standardized and reproducible way to insert gaps in one of the sequences in order to optimize alignment and to achieve a more meaningful comparison between them. Particularly in proteins, similarity is greater than identity in that conservative substitutions, for example, valine for leucine or isoleucine, are counted in calculating the reported percentage. Substitutions which are considered to be conservative are well known in the art.

"Substrate" refers to any rigid or semi-rigid support to which cDNAs or proteins are bound and includes membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, capillaries or other tubing, plates, polymers, and microparticles with a variety of surface forms including wells, trenches, pins, channels and pores.

"Variant" refers to molecules that are recognized variations of a cDNA or a protein encoded by the cDNA. Splice variants may be determined by BLAST score, wherein the score is at least 100, and most preferably at least 400. Allelic variants have a high percent identity to the cDNAs and may differ by about three bases per hundred bases. "Single nucleotide polymorphism" (SNP) refers to a

WO 02/072596

PCT/US02/07053

change in a single base as a result of a substitution, insertion or deletion. The change may be conservative (purine for purine) or non-conservative (purine to pyrimidine) and may or may not result in a change in an encoded amino acid or its secondary, tertiary, or quaternary structure.

THE INVENTION

5 The invention is based on the discovery of a cDNA which encodes STEAPRP and on the use of the cDNA, or fragments thereof, and protein, or portions thereof, directly or as compositions in the characterization, diagnosis, and treatment of prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer.

Nucleic acids encoding the STEAPRP of the present invention were first identified in Incyte
10 Clone 7100809 from the brain dentate nucleus cDNA library (BRAWTDR02) using a computer search for nucleotide and/or amino acid sequence alignments. SEQ ID NO:2 (7492448CB1) was derived from the following overlapping and/or extended nucleic acid sequences (SEQ ID NO:3-9): Incyte Clones 7100809H1 (BRAWTDR02), 6912820J1 (PITUDIR01), 4647117F6 (PROSTUT20), 7004364H1 (COLNFEC01), 70351677D1 (SG0000177), 4108079H1 (PROSBPT07), and 4669848H1
15 (SINTNOT24). Tables 1 shows expression of the transcript across the tissue categories, and the highest abundance of the transcript is found in male reproductive tissues (42%). STEAPRP is expressed exclusively in prostate tissue in this category. Table 2 shows expression of the transcript in prostate tissues, particularly in tissues from patients with adenofibromatous hyperplasia, prostate intraepithelial neoplasia, and adenocarcinoma. STEAPRP is expressed in prostate tissue libraries
20 (PROSNOT19, PROSDIT01, PROSNOT20, and PROSNOT06) from patients with adenofibromatous hyperplasia, a prostate tissue library (PROETMP06) from a patient with intraepithelial neoplasia, and prostate tissue libraries (PROSTUT18, PROSTUS20, PROSTUT04, PROSTUT21, PROSTUS19, and PROSTUT12) from patients with adenocarcinoma. Table 3 shows the differential expression of STEAPRP in human LNCaP prostate carcinoma cells compared to PrEC nontumorigenic prostate
25 epithelial cells. Cells were grown under different conditions in the experiments. Starved cells were grown in basal media in the absence of growth factors and hormones. Rich media contained growth factors and nutrients to promote growth. STEAPRP shows increased expression in LNCaP carcinoma cells relative to PrEC under all growth conditions. The transcript is therefore useful in diagnostic assays for prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer. A
30 fragment of the cDNA from about nucleotide 1 to about nucleotide 50 is also useful in diagnostic assays.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1 as shown in Figures 1A, 1B, 1C, 1D, 1E, 1F, and 1G. STEAPRP is 490 amino acids in length and has one potential N-glycosylation site at N256; one potential cyclic AMP-
35 or cyclic GMP-dependent protein kinase phosphorylation site at T32; six potential casein kinase II

WO 02/072596

PCT/US02/07053

phosphorylation sites at S12, T77, S100, S128, S197, and S348; five potential protein kinase C phosphorylation sites at S9, T46, S197, S328, and S455; and one potential tyrosine kinase phosphorylation site at Y423. PFAM analysis indicates that the region of STEAPRP from T32 to L136 is similar to a KTN NAD-binding domain. KTN NAD-binding domains are found in a variety of proteins, including potassium channels, phosphoesterases, and various transporters. BLOCKS analysis indicates that the region of STEAPRP from G34 to K64 is similar to bacterial-type phytoene dehydrogenase, the region from T32 to V56 is similar to pyridine nucleotide-disulfide class II oxidoreductases, and the region from T32 to F70 is similar to 6-phosphogluconate dehydrogenase. PRINTS analysis indicates that the region of STEAPRP from V317 to Y331 is similar to a phthalate dioxxygenase reductase family signature and the region from V33 to I47 is similar to an adrenodoxin reductase family signature. The presence of these motifs indicates a possible function for STEAPRP in oxido-reductase reactions. Hidden Markov Model analysis of STEAPRP indicates the presence of six transmembrane regions from T210 to P238, from E253 to Q281, from C301 to S328, from M359 to I379, from F391 to L411, and from F426 to I454; and the presence of a signal peptide region from M359 to N387. As shown in Figures 2A, 2B and 2C, STEAPRP has chemical and structural similarity with human STEAP (g6572948; SEQ ID NO:11). In particular, STEAPRP and STEAP share about 43% identity and the six predicted transmembrane regions. Useful antigenic epitopes extend from about G59 to about D75, from about D234 to about K249, and from about S455 to about T478; and a biologically active portion of STEAPRP extends from about T32 to about L136. An antibody which specifically binds STEAPRP is useful in a diagnostic assay to identify prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer.

Mammalian variants of the cDNA encoding STEAPRP were identified using BLAST2 with default parameters and the ZOOSEQ databases (Incyte Genomics). These preferred variants have about 85% identity as shown in the table below. The first column shows the SEQ ID for the human cDNA (SEQ ID_h); the second column, the SEQ ID for the variant cDNAs (SEQ ID_{var}); the third column, the clone number for the variant cDNAs (Clone_{var}); the fourth column, the library name; the fifth column, the alignment of the variant cDNA to the human cDNA; and the sixth column, the percent identity to the human cDNA.

SEQ ID _h	SEQ ID _{var}	Clone _{var}	Library Name	N _h Alignment	Identity
2	10	702819778T1	RATSNON03	285-607	85%

The cDNA, SEQ ID NO:10 is particularly useful for producing transgenic cell lines or organisms.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of cDNAs encoding STEAPRP, some bearing minimal similarity to the cDNAs of

WO 02/072596

PCT/US02/07053

any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of cDNA 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 encoding naturally occurring STEAPRP, and all such variations are to be considered as being specifically disclosed.

The cDNAs of SEQ ID NOs:2-10 may be used in hybridization, amplification, and screening technologies to identify and distinguish among SEQ ID NO:2 and related molecules in a sample. The mammalian cDNAs may be used to produce transgenic cell lines or organisms which are model systems for human prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer and upon which the toxicity and efficacy of potential therapeutic treatments may be tested. Toxicology studies, clinical trials, and subject/patient treatment profiles may be performed and monitored using the cDNAs, proteins, antibodies and molecules and compounds identified using the cDNAs and proteins of the present invention.

Characterization and Use of the Invention

15 cDNA libraries

In a particular embodiment disclosed herein, mRNA is isolated from mammalian cells and tissues using methods which are well known to those skilled in the art and used to prepare the cDNA libraries. The Incyte cDNAs were isolated from mammalian cDNA libraries prepared as described in the EXAMPLES. The consensus sequences are chemically and/or electronically assembled from fragments including Incyte cDNAs and extension and/or shotgun sequences using computer programs such as PHRAP (P Green, University of Washington, Seattle WA), and AUTOASSEMBLER application (Applied Biosystems, Foster City CA). After verification of the 5' and 3' sequence, at least one representative cDNA which encodes STEAPRP is designated a reagent.

Sequencing

25 Methods for sequencing nucleic acids are well known in the art and may be used to practice any of the embodiments of the invention. These methods employ enzymes such as the Klenow fragment of DNA polymerase I, SEQUENASE, Taq DNA polymerase and thermostable T7 DNA polymerase (Amersham Pharmacia Biotech (APB), 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 system (Hamilton, Reno NV) and the DNA ENGINE thermal cycler (MJ Research, Watertown MA). Machines commonly used for sequencing include the ABI PRISM 3700, 377 or 373 DNA sequencing systems (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (APB), and the like. The sequences may be analyzed using a variety of algorithms well known 35 in the art and described in Ausubel *et al.* (1997; *Short Protocols in Molecular Biology*, John Wiley &

WO 02/072596

PCT/US02/07053

Sons, New York NY, unit 7.7) and in Meyers (1995; Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

Shotgun sequencing may also be used to complete the sequence of a particular cloned insert of interest. Shotgun strategy involves randomly breaking the original insert into segments of various sizes and cloning these fragments into vectors. The fragments are sequenced and reassembled using overlapping ends until the entire sequence of the original insert is known. Shotgun sequencing methods are well known in the art and use thermostable DNA polymerases, heat-labile DNA polymerases, and primers chosen from representative regions flanking the cDNAs of interest. Incomplete assembled sequences are inspected for identity using various algorithms or programs such as CONSED (Gordon (1998) *Genome Res* 8:195-202) which are well known in the art. Contaminating sequences, including vector or chimeric sequences, or deleted sequences can be removed or restored, respectively, organizing the incomplete assembled sequences into finished sequences.

Extension of a Nucleic Acid Sequence

The sequences of the invention may be extended using various PCR-based methods known in the art. For example, the XL-PCR kit (Applied Biosystems), nested primers, and commercially available cDNA or genomic DNA libraries may be used to extend the nucleic acid sequence. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO primer analysis software (Molecular Biology Insights, Cascade CO) to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to a target molecule at temperatures from about 55C to about 68C. When extending a sequence to recover regulatory elements, it is preferable to use genomic, rather than cDNA libraries.

Hybridization

The cDNA and fragments thereof can be used in hybridization technologies for various purposes. A probe may be designed or derived from unique regions such as the 5' regulatory region or from a nonconserved region (i.e., 5' or 3' of the nucleotides encoding the conserved catalytic domain of the protein) and used in protocols to identify naturally occurring molecules encoding the STEAPRP, allelic variants, or related molecules. The probe may be DNA or RNA, may be single-stranded, and should have at least 50% sequence identity to any of the nucleic acid sequences, SEQ ID NOs:2-10. Hybridization probes may be produced using oligolabeling, nick translation, end-labeling, or PCR amplification in the presence of a reporter molecule. A vector containing the cDNA or a fragment thereof may be used to produce an mRNA probe *in vitro* by addition of an RNA polymerase and labeled nucleotides. These procedures may be conducted using commercially available kits such as those provided by APB.

The stringency of hybridization is determined by G+C content of the probe, salt concentration,

WO 02/072596

PCT/US02/07053

and temperature. In particular, stringency can be increased by reducing the concentration of salt or raising the hybridization temperature. Hybridization can be performed at low stringency with buffers, such as 5xSSC with 1% sodium dodecyl sulfate (SDS) at 60C, which permits the formation of a hybridization complex between nucleic acid sequences that contain some mismatches. Subsequent washes are performed at higher stringency with buffers such as 0.2xSSC with 0.1% SDS at either 45C (medium stringency) or 68C (high stringency). At high stringency, hybridization complexes will remain stable only where the nucleic acids are completely complementary. In some membrane-based hybridizations, preferably 35% or most preferably 50%, formamide can be added to the hybridization solution to reduce the temperature at which hybridization is performed, and background signals can be reduced by the use of detergents such as Sarkosyl or TRITON X-100 (Sigma-Aldrich, St. Louis MO) and a blocking agent such as denatured salmon sperm DNA. Selection of components and conditions for hybridization are well known to those skilled in the art and are reviewed in Ausubel (*supra*) and Sambrook *et al.* (1989) *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor Press, Plainview NY.

15 Arrays may be prepared and analyzed using methods well known in the art. Oligonucleotides or cDNAs may be used as hybridization probes or targets to monitor the expression level of large numbers of genes simultaneously or to identify genetic variants, mutations, and single nucleotide polymorphisms. Arrays may be used to determine gene function; to understand the genetic basis of a condition, disease, or disorder; to diagnose a condition, disease, or disorder; and to develop and monitor the activities of therapeutic agents. (See, e.g., Brennan *et al.* (1995) USPN 5,474,796; Schena *et al.* (1996) *Proc Natl Acad Sci* 93:10614-10619; Heller *et al.* (1997) *Proc Natl Acad Sci* 94:2150-2155; and Heller *et al.* (1997) USPN 5,605,662.)

Hybridization probes are also useful in mapping the naturally occurring genomic sequence.

The probes may be hybridized to a particular chromosome, a specific region of a chromosome, or an artificial chromosome construction. Such constructions include human artificial chromosomes (HAC), yeast artificial chromosomes (YAC), bacterial artificial chromosomes (BAC), bacterial P1 constructions, or the cDNAs of libraries made from single chromosomes.

Expression

Any one of a multitude of cDNAs encoding STEAPRP may be cloned into a vector and used to express the protein, or portions thereof, in host cells. The nucleic acid sequence can be engineered by such methods as DNA shuffling (USPN 5,830,721) and site-directed mutagenesis to create new restriction sites, alter glycosylation patterns, change codon preference to increase expression in a particular host, produce splice variants, extend half-life, and the like. The expression vector may contain transcriptional and translational control elements (promoters, enhancers, specific initiation signals, and polyadenylated 3' sequence) from various sources which have been selected for their

WO 02/072596

PCT/US02/07053

efficiency in a particular host. The vector, cDNA, and regulatory elements are combined using in vitro recombinant DNA techniques, synthetic techniques, and/or in vivo genetic recombination techniques well known in the art and described in Sambrook (supra, ch. 4, 8, 16 and 17).

A variety of host systems may be transformed with an expression vector. These include, but are not limited to, bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems transformed with baculovirus expression vectors; plant cell systems transformed with expression vectors containing viral and/or bacterial elements, or animal cell systems (Ausubel supra, unit 16). For example, an adenovirus transcription/translation complex may be utilized in mammalian cells. After sequences are ligated into the E1 or E3 region of the viral genome, the infective virus is used to transform and express the protein in host cells. The Rous sarcoma virus enhancer or SV40 or EBV-based vectors may also be used for high-level protein expression.

Routine cloning, subcloning, and propagation of nucleic acid sequences can be achieved using the multifunctional PBLUESCRIPT vector (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Introduction of a nucleic acid sequence into the multiple cloning site of these vectors disrupts the lacZ gene and allows colorimetric screening for transformed bacteria. 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.

For long term production of recombinant proteins, the vector can be stably transformed into cell lines along with a selectable or visible marker gene on the same or on a separate vector. After transformation, cells are allowed to grow for about 1 to 2 days in enriched media and then are transferred to selective media. Selectable markers, antimetabolite, antibiotic, or herbicide resistance genes, confer resistance to the relevant selective agent and allow growth and recovery of cells which successfully express the introduced sequences. Resistant clones identified either by survival on selective media or by the expression of visible markers may be propagated using culture techniques. Visible markers are also used to estimate the amount of protein expressed by the introduced genes. Verification that the host cell contains the desired cDNA is based on DNA-DNA or DNA-RNA hybridizations or PCR amplification techniques.

The host cell may be chosen for its ability to modify a recombinant protein in a desired fashion. Such modifications include acetylation, carboxylation, glycosylation, phosphorylation, lipidation, acylation and the like. Post-translational processing which cleaves a "prepro" form may also be used to specify protein targeting, folding, and/or activity. Different host cells available from the ATCC (Manassas VA) which have specific cellular machinery and characteristic mechanisms for post-translational activities may be chosen to ensure the correct modification and processing of the recombinant protein.

WO 02/072596

PCT/US02/07053

Recovery of Proteins from Cell Culture

Heterologous moieties engineered into a vector for ease of purification include glutathione S-transferase (GST), 6xHis, FLAG, MYC, and the like. GST and 6-His are purified using commercially available affinity matrices such as immobilized glutathione and metal-chelate resins, respectively.

5 FLAG and MYC are purified using commercially available monoclonal and polyclonal antibodies.

For ease of separation following purification, a sequence encoding a proteolytic cleavage site may be part of the vector located between the protein and the heterologous moiety. Methods for recombinant protein expression and purification are discussed in Ausubel (supra, unit 16) and are commercially available.

10 Chemical Synthesis of Peptides

Proteins or portions thereof may be produced not only by recombinant methods, but also by using chemical methods well known in the art. Solid phase peptide synthesis may be carried out in a batchwise or continuous flow process which sequentially adds α -amino- and side chain-protected amino acid residues to an insoluble polymeric support via a linker group. A linker group such as

15 methylamine-derivatized polyethylene glycol is attached to poly(styrene-co-divinylbenzene) to form the support resin. The amino acid residues are N- α -protected by acid labile Boc (t-butyloxycarbonyl) or base-labile Fmoc (9-fluorenylmethoxycarbonyl). The carboxyl group of the protected amino acid is coupled to the amine of the linker group to anchor the residue to the solid phase support resin.

Trifluoroacetic acid or piperidine are used to remove the protecting group in the case of Boc or Fmoc, respectively. Each additional amino acid is added to the anchored residue using a coupling agent or pre-activated amino acid derivative, and the resin is washed. The full length peptide is synthesized by sequential deprotection, coupling of derivitized amino acids, and washing with dichloromethane and/or N, N-dimethylformamide. The peptide is cleaved between the peptide carboxy terminus and the linker group to yield a peptide acid or amide. (Novabiochem 1997/98 Catalog and Peptide

25 Synthesis Handbook, San Diego CA pp. S1-S20). Automated synthesis may also be carried out on machines such as the ABI 431A peptide synthesizer (Applied Biosystems). A protein or portion thereof may be substantially purified by preparative high performance liquid chromatography and its composition confirmed by amino acid analysis or by sequencing (Creighton (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY).

30 Preparation and Screening of Antibodies

Various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with STEAPRP or any portion thereof. Adjuvants such as Freund's, mineral gels, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemacyanin (KLH), and dinitrophenol may be used to increase immunological

35 response. The oligopeptide, peptide, or portion of protein used to induce antibodies should consist of

WO 02/072596

PCT/US02/07053

at least about five amino acids, more preferably ten amino acids, which are identical to a portion of the natural protein. Oligopeptides may be fused with proteins such as KLH in order to produce antibodies to the chimeric molecule.

Monoclonal antibodies may be prepared using any technique which provides for the
5 production of antibodies 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 *et al.* (1975) *Nature* 256:495-497; Kozbor *et al.* (1985) *J. Immunol Methods* 81:31-42; Cote *et al.* (1983) *Proc Natl Acad Sci* 80:2026-2030; and Cole *et al.* (1984) *Mol Cell Biol* 62:109-120.)

10 Alternatively, techniques described for antibody production may be adapted, using methods known in the art, to produce epitope-specific, single chain antibodies. Antibody fragments which contain specific binding sites for epitopes of the protein 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
15 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 *et al.* (1989) *Science* 246:1275-1281.)

The STEAPRP or a portion thereof may be used in screening assays of phagemid or B-lymphocyte immunoglobulin libraries to identify antibodies having the desired specificity. Numerous
20 protocols for competitive binding or immunoassays 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 the protein and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes is preferred, but a competitive binding assay may also be employed (Pound (1998)
25 *Immunochemical Protocols*, Humana Press, Totowa NJ).

Labeling of Molecules for Assay

A wide variety of reporter molecules and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid, amino acid, and antibody assays. Synthesis of labeled molecules may be achieved using commercially available kits (Promega, Madison WI) for
30 incorporation of a labeled nucleotide such as ³²P-dCTP (APB), Cy3-dCTP or Cy5-dCTP (Operon Technologies, Alameda CA), or amino acid such as ³⁵S-methionine (APB). Nucleotides and amino acids may be directly labeled with a variety of substances including fluorescent, chemiluminescent, or chromogenic agents, and the like, by chemical conjugation to amines, thiols and other groups present in the molecules using reagents such as BIODIPY or FITC (Molecular Probes, Eugene OR).

35 **DIAGNOSTICS**

WO 02/072596

PCT/US02/07053

The cDNAs, fragments, oligonucleotides, complementary RNA and DNA molecules, and PNAs and may be used to detect and quantify differential gene expression for diagnosis of a disorder. Similarly antibodies which specifically bind STEAPRP may be used to quantitate the protein. Disorders associated with differential expression include prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer. The diagnostic assay may use hybridization or amplification technology to compare gene expression in a biological sample from a patient to standard samples in order to detect differential gene expression. Qualitative or quantitative methods for this comparison are well known in the art.

For example, the cDNA or probe may be labeled by standard methods and added to a biological sample from a patient under conditions for the formation of hybridization complexes. After an incubation period, the sample is washed and the amount of label (or signal) associated with hybridization complexes, is quantified and compared with a standard value. If complex formation in the patient sample is significantly altered (higher or lower) in comparison to either a normal or disease standard, then differential expression indicates the presence of a disorder.

In order to provide standards for establishing differential expression, normal and disease expression profiles are established. This is accomplished by combining a sample taken from normal subjects, either animal or human, with a cDNA under conditions for hybridization to occur. Standard hybridization complexes may be quantified by comparing the values obtained using normal subjects with values from an experiment in which a known amount of a purified sequence is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who were diagnosed with a particular condition, disease, or disorder. Deviation from standard values toward those associated with a particular disorder is used to diagnose that disorder.

Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies or in clinical trials or to monitor the treatment of an individual patient. Once the presence of a condition is established and a treatment protocol is initiated, diagnostic 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 a 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.

Immunological Methods

Detection and quantification of a protein 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 is preferred, but a competitive binding assay may be employed. (See, e.g., Coligan et al. (1997) Current Protocols in Immunology, Wiley-Interscience, New York NY; and

WO 02/072596

PCT/US02/07053

Pound, *supra.*)**THERAPEUTICS**

Chemical and structural similarity, in particular the six transmembrane domains, exists between regions of STEAPRP (SEQ ID NO:1) and human STEAP (g6572948; SEQ ID NO:11) as shown in Figures 2A, 2B, and 2C. In addition, differential expression is highly associated with LNCaP prostate carcinoma cells and prostate tissues and with prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer as shown in Tables 1-3. STEAPRP clearly plays a role in prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer.

In the treatment of conditions associated with increased expression of the STEAPRP, it is desirable to decrease expression or protein activity. In one embodiment, the an inhibitor, antagonist, or antibody of the protein may be administered to a subject to treat a condition associated with increased expression or activity. In another embodiment, a pharmaceutical composition comprising an inhibitor, antagonist or antibody in conjunction with a pharmaceutical carrier may be administered to a subject to treat a condition associated with the increased expression or activity of the endogenous protein. In an additional embodiment, a vector expressing the complement of the cDNA or fragments thereof may be administered to a subject to treat the disorder.

In the treatment of conditions associated with decreased expression of the STEAPRP, it is desirable to increase expression or protein activity. In one embodiment, the protein, an agonist, or enhancer may be administered to a subject to treat a condition associated with decreased expression or activity. In another embodiment, a pharmaceutical composition comprising the protein, an agonist or enhancer in conjunction with a pharmaceutical carrier may be administered to a subject to treat a condition associated with the decreased expression or activity of the endogenous protein. In an additional embodiment, a vector expressing cDNA may be administered to a subject to treat the disorder.

Any of the cDNAs, complementary molecules, or fragments thereof, proteins or portions thereof, vectors delivering these nucleic acid molecules or expressing the proteins, and their ligands may be administered in combination with other therapeutic agents. Selection of the agents for use in combination therapy may be made by one of ordinary skill in the art according to conventional pharmaceutical principles. A combination of therapeutic agents may act synergistically to affect treatment of a particular disorder at a lower dosage of each agent.

Modification of Gene Expression Using Nucleic Acids

Gene expression may be modified by designing complementary or antisense molecules (DNA, RNA, or PNA) to the control, 5', 3', or other regulatory regions of the gene encoding STEAPRP. Oligonucleotides designed to inhibit transcription initiation are preferred. Similarly, inhibition can be achieved using triple helix base-pairing which inhibits the binding of polymerases, transcription

WO 02/072596

PCT/US02/07053

factors, or regulatory molecules (Gee et al. in: Huber and Carr (1994) Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary molecule may also be designed to block translation by preventing binding between ribosomes and mRNA. In one alternative, a library or plurality of cDNAs may be screened to identify those which specifically bind a
 5 regulatory, nontranslated sequence.

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 at sites such as GUA, GUU, and GUC. Once such sites are identified, an oligonucleotide with the same sequence may be
 10 evaluated for secondary structural features which would render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing their hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary nucleic acids and ribozymes of the invention may be prepared via recombinant expression, in vitro or in vivo, or using solid phase phosphoramidite chemical synthesis.
 15 In addition, RNA molecules may be modified to increase intracellular stability and half-life by addition of flanking sequences at the 5' and/or 3' ends of the molecule or by the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. Modification is inherent in the production of PNAs and can be extended to other nucleic acid molecules. Either the inclusion of nontraditional bases such as inosine, queosine, and
 20 wybutosine, and or the modification of adenine, cytidine, guanine, thymine, and uridine with acetyl-, methyl-, thio- groups renders the molecule less available to endogenous endonucleases.

Screening and Purification Assays

The cDNA encoding STEAPRP may be used to screen a library of molecules or compounds for specific binding affinity. The libraries may be aptamers, DNA molecules, RNA molecules, PNAs,
 25 peptides, proteins such as transcription factors, enhancers, repressors, and other ligands which regulate the activity, replication, transcription, or translation of the endogenous gene. The assay involves combining a polynucleotide with a library of molecules under conditions allowing specific binding, and detecting specific binding to identify at least one molecule which specifically binds the single-stranded or double-stranded molecule.

30 In one embodiment, the cDNA of the invention may be incubated with a plurality of purified molecules or compounds and binding activity determined by methods well known in the art, e.g., a gel-retardation assay (USPN 6,010,849) or a reticulocyte lysate transcriptional assay. In another embodiment, the cDNA may be incubated with nuclear extracts from biopsied and/or cultured cells and tissues. Specific binding between the cDNA and a molecule or compound in the nuclear extract is
 35 initially determined by gel shift assay and may be later confirmed by recovering and raising antibodies

WO 02/072596

PCT/US02/07053

against that molecule or compound. When these antibodies are added into the assay, they cause a supershift in the gel-retardation assay.

In another embodiment, the cDNA may be used to purify a molecule or compound using affinity chromatography methods well known in the art. In one embodiment, the cDNA is chemically
5 reacted with cyanogen bromide groups on a polymeric resin or gel. Then a sample is passed over and reacts with or binds to the cDNA. The molecule or compound which is bound to the cDNA may be released from the cDNA by increasing the salt concentration of the flow-through medium and collected.

In a further embodiment, the protein or a portion thereof may be used to purify a ligand from a
10 sample. A method for using a protein or a portion thereof to purify a ligand would involve combining the protein or a portion thereof with a sample under conditions to allow specific binding, detecting specific binding between the protein and ligand, recovering the bound protein, and using an appropriate chaotropic agent to separate the protein from the purified ligand.

In a preferred embodiment, STEAPRP may be used to screen a plurality of molecules or
15 compounds in any of a variety of screening assays. The portion of the protein employed in such screening may be free in solution, affixed to an abiotic or biotic substrate (e.g. borne on a cell surface), or located intracellularly. For example, in one method, viable or fixed prokaryotic host cells that are stably transformed with recombinant nucleic acids that have expressed and positioned a peptide on their cell surface can be used in screening assays. The cells are screened against a plurality
20 or libraries of ligands, and the specificity of binding or formation of complexes between the expressed protein and the ligand may be measured. Specific binding between the protein and molecule may be measured. Depending on the particular kind of library being screened, the assay may be used to identify DNA molecules, RNA molecules, peptide nucleic acids, peptides, proteins, mimetics, agonists, antagonists, antibodies, immunoglobulins, inhibitors, and drugs or any other ligand, which
25 specifically binds the protein.

In one aspect, this invention contemplates a method for high throughput screening using very small assay volumes and very small amounts of test compound as described in USPN 5,876,946, incorporated herein by reference. This method is used to screen large numbers of molecules and compounds via specific binding. In another aspect, this invention also contemplates the use of
30 competitive drug screening assays in which neutralizing antibodies capable of binding the protein specifically compete with a test compound capable of binding to the protein. Molecules or compounds identified by screening may be used in a mammalian model system to evaluate their toxicity, diagnostic, or therapeutic potential.

Pharmacology

35 Pharmaceutical compositions are those substances wherein the active ingredients are

WO 02/072596

PCT/US02/07053

contained in an effective amount to achieve a desired and intended purpose. The determination of an effective dose is well within the capability of those skilled in the art. For any compound, the therapeutically effective dose may be estimated initially either in cell culture assays or in animal models. The animal model is also used to achieve a desirable concentration range and route of administration. Such information may then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of protein or inhibitor which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity of such agents may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED_{50} (the dose therapeutically effective in 50% of the population) and LD_{50} (the dose lethal to 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index, and it may be expressed as the ratio, LD_{50}/ED_{50} . Pharmaceutical compositions which exhibit large therapeutic indexes are preferred. The data obtained from cell culture assays and animal studies are used in formulating a range of dosage for human use.

15 Model Systems

Animal models may be used as bioassays where they exhibit a phenotypic response similar to that of humans and where exposure conditions are relevant to human exposures. Mammals are the most common models, and most infectious agent, cancer, drug, and toxicity studies are performed on rodents such as rats or mice because of low cost, availability, lifespan, reproductive potential, and abundant reference literature. Inbred and outbred rodent strains provide a convenient model for investigation of the physiological consequences of under- or over-expression of genes of interest and for the development of methods for diagnosis and treatment of diseases. A mammal inbred to over-express a particular gene (for example, secreted in milk) may also serve as a convenient source of the protein expressed by that gene.

25 Toxicology

Toxicology is the study of the effects of agents on living systems. The majority of toxicity studies are performed on rats or mice. Observation of qualitative and quantitative changes in physiology, behavior, homeostatic processes, and lethality in the rats or mice are used to generate a toxicity profile and to assess potential consequences on human health following exposure to the agent.

Genetic toxicology identifies and analyzes the effect of an agent on the rate of endogenous, spontaneous, and induced genetic mutations. Genotoxic agents usually have common chemical or physical properties that facilitate interaction with nucleic acids and are most harmful when chromosomal aberrations are transmitted to progeny. Toxicological studies may identify agents that increase the frequency of structural or functional abnormalities in the tissues of the progeny if administered to either parent before conception, to the mother during pregnancy, or to the developing

WO 02/072596

PCT/US02/07053

organism. Mice and rats are most frequently used in these tests because their short reproductive cycle allows the production of the numbers of organisms needed to satisfy statistical requirements.

Acute toxicity tests are based on a single administration of an agent to the subject to determine the symptomatology or lethality of the agent. Three experiments are conducted: 1) an initial 5 dose-range-finding experiment, 2) an experiment to narrow the range of effective doses, and 3) a final experiment for establishing the dose-response curve.

Subchronic toxicity tests are based on the repeated administration of an agent. Rat and dog are commonly used in these studies to provide data from species in different families. With the exception of carcinogenesis, there is considerable evidence that daily administration of an agent at 10 high-dose concentrations for periods of three to four months will reveal most forms of toxicity in adult animals.

Chronic toxicity tests, with a duration of a year or more, are used to demonstrate either the absence of toxicity or the carcinogenic potential of an agent. When studies are conducted on rats, a minimum of three test groups plus one control group are used, and animals are examined and 15 monitored at the outset and at intervals throughout the experiment.

Transgenic Animal Models

Transgenic rodents that over-express or under-express a gene of interest may be inbred and used to model human diseases or to test therapeutic or toxic agents. (See, e.g., USPN 5,175,383 and USPN 5,767,337.) In some cases, the introduced gene may be activated at a specific time in a specific 20 tissue type during fetal or postnatal development. Expression of the transgene is monitored by analysis of phenotype, of tissue-specific mRNA expression, or of serum and tissue protein levels in transgenic animals before, during, and after challenge with experimental drug therapies.

Embryonic Stem Cells

Embryonic (ES) stem cells isolated from rodent embryos retain the potential to form 25 embryonic tissues. When ES cells are placed inside a carrier embryo, they resume normal development and contribute to tissues of the live-born animal. ES cells are the preferred cells used in the creation of experimental knockout and knockin rodent strains. Mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and are grown under culture conditions well known in the art. Vectors used to produce a transgenic strain contain a disease gene candidate 30 and a marker gene, the latter serves to identify the presence of the introduced disease gene. The vector is transformed into ES cells by methods well known in the art, and 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.

35 ES cells derived from human blastocysts may be manipulated *in vitro* to differentiate into at

WO 02/072596

PCT/US02/07053

least eight separate cell lineages. These lineages are used to study the differentiation of various cell types and tissues in vitro, and they include endoderm, mesoderm, and ectodermal cell types which differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes.

Knockout Analysis

5 In gene knockout analysis, a region of a mammalian gene is enzymatically modified to include a non-mammalian gene such as the neomycin phosphotransferase gene (neo; Capecchi (1989) Science 244:1288-1292). The modified gene is transformed into cultured ES cells and integrates into the endogenous genome by homologous recombination. The inserted sequence disrupts transcription and translation of the endogenous gene. Transformed cells are injected into rodent blastulae, and the
10 blastulae are implanted into pseudopregnant dams. Transgenic progeny are crossbred to obtain homozygous inbred lines which lack a functional copy of the mammalian gene. In one example, the mammalian gene is a human gene.

Knockin Analysis

ES cells can be used to create knockin humanized animals (pigs) or transgenic animal models
15 (mice or rats) of human diseases. With knockin technology, a region of a human gene is injected into animal ES cells, and the human 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 the analogous human condition. These methods have been used to model several human
20 diseases.

Non-Human Primate Model

The field of animal testing deals with data and methodology from basic sciences such as physiology, genetics, chemistry, pharmacology and statistics. These data are paramount in evaluating the effects of therapeutic agents on non-human primates as they can be related to human health.
25 Monkeys are used as human surrogates in vaccine and drug evaluations, and their responses are relevant to human exposures under similar conditions. Cynomolgus and Rhesus monkeys (Macaca fascicularis and Macaca mulatta, respectively) and Common Marmosets (Callithrix jacchus) are the most common non-human primates (NHPs) used in these investigations. Since great cost is associated with developing and maintaining a colony of NHPs, early research and toxicological studies are
30 usually carried out in rodent models. In studies using behavioral measures such as drug addiction, NHPs are the first choice test animal. In addition, NHPs and individual humans exhibit differential sensitivities to many drugs and toxins and can be classified as a range of phenotypes from "extensive metabolizers" to "poor metabolizers" of these agents.

In additional embodiments, the cDNAs which encode the protein may be used in any
35 molecular biology techniques that have yet to be developed, provided the new techniques rely on

WO 02/072596

PCT/US02/07053

properties of cDNAs that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

EXAMPLES

5 The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention. The preparation of the human brain dentate nucleus (BRAWTDR02) library will be described.

I cDNA Library Construction

The BRAWTDR02 cDNA library was constructed from brain dentate nucleus tissue removed
10 from a 55-year-old Caucasian female (specimen #A98-58) who died from cholangiocarcinoma. The frozen tissue was homogenized and lysed in TRIZOL reagent (0.8 g tissue/12 ml; Life Technologies) using a POLYTRON homogenizer (Brinkmann Instruments, Westbury NJ). The lysate was centrifuged over a 5.7 M CsCl cushion using an SW28 rotor in an LS-70M ultracentrifuge (Beckman Coulter, Fullerton CA) for 18 hours at 25,000 rpm at ambient temperature. The RNA was extracted
15 with acid phenol, pH 4.7, precipitated using 0.3 M sodium acetate and 2.5 volumes of ethanol, resuspended in RNase-free water, and treated with DNase at 37C. The RNA was reextracted and precipitated as before. The mRNA was isolated with the OLIGOTEX kit (Qiagen, Chatsworth CA) and used to construct the cDNA library.

The mRNA was handled according to the recommended protocols in the SUPERScript
20 plasmid system (Life Technologies) which contains a NotI primer-adaptor designed to prime the first strand cDNA synthesis at the poly(A) tail of mRNAs. Double stranded cDNA was blunted, ligated to EcoRI adaptors and digested with NotI (New England Biolabs, Beverly MA). The cDNAs were fractionated on a SEPHAROSE CL4B column (APB), and those cDNAs exceeding 400 bp were ligated into pcDNA2.1 plasmid (Invitrogen, Carlsbad CA). The plasmid pcDNA2.1 was subsequently
25 transformed into DH5α competent cells (Life Technologies).

II Isolation and Sequencing of cDNA Clones

Plasmid DNA was released from the cells and purified using either the MINIPREP kit (Edge Biosystems, Gaithersburg MD) or the REAL PREP 96 plasmid kit (Qiagen). A kit consists of a 96-well block with reagents for 960 purifications. The recommended protocol was employed except for
30 the following changes: 1) the bacteria were cultured in 1 ml of sterile TERRIFIC BROTH (APB) with carbenicillin at 25 mg/l and glycerol at 0.4%; 2) after inoculation, the cells were cultured for 19 hours and then lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were transferred to a 96-well block for storage at 4C.

35 The cDNAs were prepared for sequencing using the MICROLAB 2200 system (Hamilton) in

WO 02/072596

PCT/US02/07053

combination with the DNA ENGINE thermal cyclers (MJ Research). The cDNAs were sequenced by the method of Sanger and Coulson (1975; J Mol Biol 94:441-448) using an ABI PRISM 377 sequencing system (Applied Biosystems) or the MEGABACE 1000 DNA sequencing system (APB). Most of the isolates were sequenced according to standard ABI protocols and kits (Applied Biosystems) with solution volumes of 0.25x-1.0x concentrations. In the alternative, cDNAs were sequenced using solutions and dyes from APB.

III Extension of cDNA Sequences

The cDNAs were extended using the cDNA clone and oligonucleotide primers. One primer was synthesized to initiate 5' extension of the known fragment, and the other, to initiate 3' extension of the known fragment. The initial primers were designed using commercially available primer analysis software 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 68C to about 72C. Any stretch of nucleotides that would result in hairpin structures and primer-primer dimerizations was avoided.

Selected cDNA libraries were used as templates to extend the sequence. If more than one extension was necessary, additional or nested sets of primers were designed. Preferred libraries have been size-selected to include larger cDNAs and random primed to contain more sequences with 5' or upstream regions of genes. Genomic libraries are used to obtain regulatory elements, especially extension into the 5' promoter binding region.

High fidelity amplification was obtained by PCR using methods such as that taught in USPN 5,932,451. PCR was performed in 96-well plates using the DNA ENGINE thermal cycler (MJ Research). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+} , $(NH_4)_2SO_4$, and β -mercaptoethanol, Taq DNA polymerase (APB), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B (Incyte Genomics): Step 1: 94C, three min; Step 2: 94C, 15 sec; Step 3: 60C, one min; Step 4: 68C, two min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68C, five min; Step 7: storage at 4C. In the alternative, the parameters for primer pair T7 and SK+ (Stratagene) were as follows: Step 1: 94C, three min; Step 2: 94C, 15 sec; Step 3: 57C, one min; Step 4: 68C, two min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68C, five min; Step 7: storage at 4C.

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

WO 02/072596

PCT/US02/07053

The extended clones were desalted, concentrated, transferred to 384-well plates, digested with CviII cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC18 vector (APB). For shotgun sequences, the digested nucleotide sequences were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and the agar was digested with AGARACE enzyme (Promega). Extended clones were religated using T4 DNA ligase (New England Biolabs) into pUC18 vector (APB), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into *E. coli* competent cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37C in 384-well plates in LB/2x carbenicillin liquid media.

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

IV Homology Searching of cDNA Clones and Their Deduced Proteins

20 The cDNAs of the Sequence Listing or their deduced amino acid sequences were used to query databases such as GenBank, SwissProt, BLOCKS, and the like. These databases that contain previously identified and annotated sequences or domains were searched using BLAST or BLAST2 to produce alignments and to determine which sequences were exact matches or homologs. The alignments were to sequences of prokaryotic (bacterial) or eukaryotic (animal, fungal, or plant) origin. 25 Alternatively, algorithms such as the one described in Smith and Smith (1992, Protein Engineering 5:35-51) could have been used to deal with primary sequence patterns and secondary structure gap penalties. All of the sequences disclosed in this application have lengths of at least 49 nucleotides, and no more than 12% uncalled bases (where N is recorded rather than A, C, G, or T).

As detailed in Karlin (supra), BLAST matches between a query sequence and a database 30 sequence were evaluated statistically and only reported when they satisfied the threshold of 10^{-25} for nucleotides and 10^{-14} for peptides. Homology was also evaluated by product score calculated as follows: the % nucleotide or amino acid identity [between the query and reference sequences] in BLAST is multiplied by the % maximum possible BLAST score [based on the lengths of query and reference sequences] and then divided by 100. In comparison with hybridization procedures used in 35 the laboratory, the stringency for an exact match was set from a lower limit of about 40 (with 1-2%

WO 02/072596

PCT/US02/07053

error due to uncalled bases) to a 100% match of about 70.

The BLAST software suite (NCBI, Bethesda MD; <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>), includes various sequence analysis programs including "blastn" that is used to align nucleotide sequences and BLAST2 that is used for direct pairwise comparison of either nucleotide or amino acid sequences. BLAST programs are commonly used with gap and other parameters set to default settings, e.g.: 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; and Filter: on. Identity is measured over the entire length of a sequence. Brenner *et al.* (1998; *Proc Natl Acad Sci* 95:6073-6078, incorporated herein by reference) analyzed BLAST for its ability to identify structural homologs by sequence identity and found 30% identity is a reliable threshold for sequence alignments of at least 150 residues and 40%, for alignments of at least 70 residues.

The cDNAs of this application were compared with assembled consensus sequences or templates found in the LIFESEQ GOLD database (Incyte Genomics). Component sequences from cDNA, extension, full length, and shotgun sequencing projects were subjected to PHRED analysis and assigned a quality score. All sequences with an acceptable quality score were subjected to various pre-processing and editing pathways to remove low quality 3' ends, vector and linker sequences, polyA tails, Alu repeats, mitochondrial and ribosomal sequences, and bacterial contamination sequences. Edited sequences had to be at least 50 bp in length, and low-information sequences and repetitive elements such as dinucleotide repeats, Alu repeats, and the like, were replaced by "Ns" or masked.

Edited sequences were subjected to assembly procedures in which the sequences were assigned to gene bins. Each sequence could only belong to one bin, and sequences in each bin were assembled to produce a template. Newly sequenced components were added to existing bins using BLAST and CROSSMATCH. To be added to a bin, the component sequences had to have a BLAST quality score greater than or equal to 150 and an alignment of at least 82% local identity. The sequences in each bin were assembled using PHRAP. Bins with several overlapping component sequences were assembled using DEEP PHRAP. The orientation of each template was determined based on the number and orientation of its component sequences.

Bins were compared to one another, and those having local similarity of at least 82% were combined and reassembled. Bins having templates with less than 95% local identity were split. Templates were subjected to analysis by STITCHER/EXON MAPPER algorithms that determine the probabilities of the presence of splice variants, alternatively spliced exons, splice junctions, differential expression of alternative spliced genes across tissue types or disease states, and the like. Assembly procedures were repeated periodically, and templates were annotated using BLAST against

WO 02/072596

PCT/US02/07053

GenBank databases such as GBpri. An exact match was defined as having from 95% local identity over 200 base pairs through 100% local identity over 100 base pairs and a homolog match as having an E-value (or probability score) of $\leq 1 \times 10^{-8}$. The templates were also subjected to frameshift FASTx against GENPEPT, and homolog match was defined as having an E-value of $\leq 1 \times 10^{-8}$. Template analysis and assembly was described in USSN 09/276,534, filed March 25, 1999.

Following assembly, templates were subjected to BLAST, motif, and other functional analyses and categorized in protein hierarchies using methods described in USSN 08/812,290 and USSN 08/811,758, both filed March 6, 1997; in USSN 08/947,845, filed October 9, 1997; and in USSN 09/034,807, filed March 4, 1998. Then templates were analyzed by translating each template in all three forward reading frames and searching each translation against the PFAM database of hidden Markov model-based protein families and domains using the HMMER software package (Washington University School of Medicine, St. Louis MO; <http://pfam.wustl.edu/>). The cDNA was further analyzed using MACDNASIS PRO software (Hitachi Software Engineering), and LASERGENE software (DNASTAR) and queried against public databases such as the GenBank rodent, mammalian, vertebrate, prokaryote, and eukaryote databases, SwissProt, BLOCKS, PRINTS, PFAM, and Prosite.

V Chromosome Mapping

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 Génethon are used to determine if any of the cDNAs presented in the Sequence Listing have been mapped. Any of the fragments of the cDNA encoding STEAPRP that have been mapped result in the assignment of all related regulatory and coding sequences mapping to the same location. The genetic map locations are described as ranges, or intervals, of human chromosomes. The map position of an interval, in cM (which is roughly equivalent to 1 megabase of human DNA), is measured relative to the terminus of the chromosomal p-arm.

VI Hybridization Technologies and Analyses

Immobilization of cDNAs on a Substrate

The cDNAs are applied to a substrate by one of the following methods. A mixture of cDNAs is fractionated by gel electrophoresis and transferred to a nylon membrane by capillary transfer. Alternatively, the cDNAs are individually ligated to a vector and inserted into bacterial host cells to form a library. The cDNAs are then arranged on a substrate by one of the following methods. In the first method, bacterial cells containing individual clones are robotically picked and arranged on a nylon membrane. The membrane is placed on LB agar containing selective agent (carbenicillin, kanamycin, ampicillin, or chloramphenicol depending on the vector used) and incubated at 37°C for 16 hr. The membrane is removed from the agar and consecutively placed colony side up in 10% SDS,

WO 02/072596

PCT/US02/07053

denaturing solution (1.5 M NaCl, 0.5 M NaOH), neutralizing solution (1.5 M NaCl, 1 M Tris, pH 8.0), and twice in 2xSSC for 10 min each. The membrane is then UV irradiated in a STRATALINKER UV-crosslinker (Stratagene).

In the second method, cDNAs are amplified from bacterial vectors by thirty cycles of PCR using primers complementary to vector sequences flanking the insert. PCR amplification increases a starting concentration of 1-2 ng nucleic acid to a final quantity greater than 5 µg. Amplified nucleic acids from about 400 bp to about 5000 bp in length are purified using SEPHACRYL-400 beads (APB). Purified nucleic acids are arranged on a nylon membrane manually or using a dot/slot blotting manifold and suction device and are immobilized by denaturation, neutralization, and UV irradiation as described above. Purified nucleic acids are robotically arranged and immobilized on polymer-coated glass slides using the procedure described in USPN 5,807,522. Polymer-coated slides are prepared by cleaning glass microscope slides (Corning, Acton MA) by ultrasound in 0.1% SDS and acetone, etching in 4% hydrofluoric acid (VWR Scientific Products, West Chester PA), coating with 0.05% aminopropyl silane (Sigma Aldrich) in 95% ethanol, and curing in a 110°C oven. The slides are washed extensively with distilled water between and after treatments. The nucleic acids are arranged on the slide and then immobilized by exposing the array to UV irradiation using a STRATALINKER UV-crosslinker (Stratagene). Arrays are then washed at room temperature in 0.2% SDS and rinsed three times in distilled water. Non-specific binding sites are blocked by incubation of arrays in 0.2% casein in phosphate buffered saline (PBS; Tropix, Bedford MA) for 30 min at 60°C; then the arrays are washed in 0.2% SDS and rinsed in distilled water as before.

Probe Preparation for Membrane Hybridization

Hybridization probes derived from the cDNAs of the Sequence Listing are employed for screening cDNAs, mRNAs, or genomic DNA in membrane-based hybridizations. Probes are prepared by diluting the cDNAs to a concentration of 40-50 ng in 45 µl TE buffer, denaturing by heating to 100°C for five min, and briefly centrifuging. The denatured cDNA is then added to a REDIPRIME tube (APB); gently mixed until blue color is evenly distributed, and briefly centrifuged. Five µl of [³²P]dCTP is added to the tube, and the contents are incubated at 37°C for 10 min. The labeling reaction is stopped by adding 5 µl of 0.2M EDTA, and probe is purified from unincorporated nucleotides using a PROBEQUANT G-50 microcolumn (APB). The purified probe is heated to 100°C for five min, snap cooled for two min on ice, and used in membrane-based hybridizations as described below.

Probe Preparation for Polymer Coated Slide Hybridization

Hybridization probes derived from mRNA isolated from samples are employed for screening cDNAs of the Sequence Listing in array-based hybridizations. Probe is prepared using the GEMbright kit (Incyte Genomics) by diluting mRNA to a concentration of 200 ng in 9 µl TE buffer and adding 5

WO 02/072596

PCT/US02/07053

1 μ l 5x buffer, 1 μ l 0.1 M DTT, 3 μ l Cy3 or Cy5 labeling mix, 1 μ l RNase inhibitor, 1 μ l reverse
 transcriptase, and 5 μ l 1x yeast control mRNAs. Yeast control mRNAs are synthesized by in vitro
 transcription from noncoding yeast genomic DNA (W. Lei, unpublished). As quantitative controls,
 one set of control mRNAs at 0.002 ng, 0.02 ng, 0.2 ng, and 2 ng are diluted into reverse transcription
 5 reaction mixture at ratios of 1:100,000, 1:10,000, 1:1000, and 1:100 (w/w) to sample mRNA
 respectively. To examine mRNA differential expression patterns, a second set of control mRNAs are
 diluted into reverse transcription reaction mixture at ratios of 1:3, 3:1, 1:10, 10:1, 1:25, and 25:1
 (w/w). The reaction mixture is mixed and incubated at 37C for two hr. The reaction mixture is then
 incubated for 20 min at 85C, and probes are purified using two successive CHROMA SPIN+TE 30
 10 columns (Clontech, Palo Alto CA). Purified probe is ethanol precipitated by diluting probe to 90 μ l in
 DEPC-treated water, adding 2 μ l 1mg/ml glycogen, 60 μ l 5 M sodium acetate, and 300 μ l 100%
 ethanol. The probe is centrifuged for 20 min at 20,800xg, and the pellet is resuspended in 12 μ l
 resuspension buffer, heated to 65C for five min, and mixed thoroughly. The probe is heated and
 mixed as before and then stored on ice. Probe is used in high density array-based hybridizations as
 15 described below.

Membrane-based Hybridization

Membranes are pre-hybridized in hybridization solution containing 1% Sarkosyl and 1x high
 phosphate buffer (0.5 M NaCl, 0.1 M Na_2HPO_4 , 5 mM EDTA, pH 7) at 55C for two hr. The probe,
 diluted in 15 ml fresh hybridization solution, is then added to the membrane. The membrane is
 20 hybridized with the probe at 55C for 16 hr. Following hybridization, the membrane is washed for 1.5
 min at 25C in 1mM Tris (pH 8.0), 1% Sarkosyl, and four times for 15 min each at 25C in 1mM Tris
 (pH 8.0). To detect hybridization complexes, XOMAT-AR film (Eastman Kodak, Rochester NY) is
 exposed to the membrane overnight at -70C, developed, and examined visually.

Polymer Coated Slide-based Hybridization

25 Probe is heated to 65C for five min, centrifuged five min at 9400 rpm in a 5415C
 microcentrifuge (Eppendorf Scientific, Westbury NY), and then 18 μ l is aliquoted onto the array
 surface and covered with a coverslip. The arrays are transferred to a waterproof chamber having a
 cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally
 by the addition of 140 μ l of 5xSSC in a corner of the chamber. The chamber containing the arrays is
 30 incubated for about 6.5 hr at 60C. The arrays are washed for 10 min at 45C in 1xSSC, 0.1% SDS, and
 three times for 10 min each at 45C in 0.1xSSC, and dried.

Hybridization reactions are performed in absolute or differential hybridization formats. In the
 absolute hybridization format, probe from one sample is hybridized to array elements, and signals are
 detected after hybridization complexes form. Signal strength correlates with probe mRNA levels in
 35 the sample. In the differential hybridization format, differential expression of a set of genes in two

WO 02/072596

PCT/US02/07053

biological samples is analyzed. Probes from the two samples are prepared and labeled with different labeling moieties. A mixture of the two labeled probes is hybridized to the array elements, and signals are examined under conditions in which the emissions from the two different labels are individually detectable. Elements on the array that are hybridized to substantially equal numbers of probes derived from both biological samples give a distinct combined fluorescence (Shalon WO95/35505).

Hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, 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, 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 with a resolution of 20 micrometers. In the differential hybridization format, the two fluorophores are sequentially excited by the laser. 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. The sensitivity of the scans is calibrated using the signal intensity generated by the yeast control mRNAs added to the probe mix. 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.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, 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 the emission spectrum for each fluorophore. 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 program (Incyte Genomics).

VII Electronic Analysis

BLAST was used to search for identical or related molecules in the GenBank or LIFESEQ databases (Incyte Genomics). The product score for human and rat sequences was calculated as follows: the BLAST score is multiplied by the % nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences), such that a 100% alignment over the length of

WO 02/072596

PCT/US02/07053

the shorter sequence gives a product score of 100. The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and with a product score of at least 70, the match will be exact. Similar or related molecules are usually identified by selecting those 5 which show product scores between 8 and 40.

Electronic northern analysis was performed at a product score of 70 and is shown in Tables 1 and 2. All sequences and cDNA libraries in the LIFESEQ database were categorized by system, organ/tissue and cell type. The categories included cardiovascular system, connective tissue, digestive system, embryonic structures, endocrine system, exocrine glands, female and male genitalia, germ 10 cells, hemic/immune system, liver, musculoskeletal system, nervous system, pancreas, respiratory system, sense organs, skin, stomatognathic system, unclassified/mixed, and the urinary tract. For each category, the number of libraries in which the sequence was expressed were counted and shown over the total number of libraries in that category. In a non-normalized library, expression levels of two or more are significant.

15 VIII Complementary Molecules

Molecules complementary to the cDNA, from about 5 (PNA) to about 5000 bp (complement of a cDNA insert), are used to detect or inhibit gene expression. Detection is described in Example VI. To inhibit transcription by preventing promoter binding, the complementary molecule is designed to bind to the most unique 5' sequence and includes nucleotides of the 5' UTR upstream of the 20 initiation codon of the open reading frame. Complementary molecules include genomic sequences (such as enhancers or introns) and are used in "triple helix" base pairing to compromise the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. To inhibit translation, a complementary molecule is designed to prevent ribosomal binding to the mRNA encoding the protein.

25 Complementary molecules are placed in expression vectors and used to transform a cell line to test efficacy; into an organ, tumor, synovial cavity, or the vascular system for transient or short term therapy; or into a stem cell, zygote, or other reproducing lineage for long term or stable gene therapy. Transient expression lasts for a month or more with a non-replicating vector and for three months or more if appropriate elements for inducing vector replication are used in the transformation/expression 30 system.

Stable transformation of appropriate dividing cells with a vector encoding the complementary molecule produces a transgenic cell line, tissue, or organism (USPN 4,736,866). Those cells that assimilate and replicate sufficient quantities of the vector to allow stable integration also produce enough complementary molecules to compromise or entirely eliminate activity of the cDNA encoding 35 the protein.

WO 02/072596

PCT/US02/07053

IX Selection of Sequences, Microarray Preparation and Use

Incyte clones represent template sequences derived from the LIFESEQ GOLD assembled human sequence database (Incyte Genomics). In cases where more than one clone was available for a particular template, the 5'-most clone in the template was used on the microarray. The HUMAN GENOME GEM series 1-3 microarrays (Incyte Genomics) contain 28,626 array elements which represent 10,068 annotated clusters and 18,558 unannotated clusters. For the UNIGEM series microarrays (Incyte Genomics), Incyte clones were mapped to non-redundant Unigene clusters (Unigene database (build 46), NCBI; Shuler (1997) J Mol Med 75:694-698), and the 5' clone with the strongest BLAST alignment (at least 90% identity and 100 bp overlap) was chosen, verified, and used in the construction of the microarray. The UNIGEM V microarray (Incyte Genomics) contains 7075 array elements which represent 4610 annotated genes and 2,184 unannotated clusters.

To construct microarrays, cDNAs were amplified from bacterial cells using primers complementary to vector sequences flanking the cDNA insert. Thirty cycles of PCR increased the initial quantity of cDNAs from 1-2 ng to a final quantity of greater than 5 μ g. Amplified cDNAs were then purified using SEPHACRYL-400 columns (APB). Purified cDNAs were immobilized on polymer-coated glass slides. Glass microscope slides (Corning, Corning NY) were cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides were etched in 4% hydrofluoric acid (VWR Scientific Products, West Chester PA), washed thoroughly in distilled water, and coated with 0.05% aminopropyl silane (Sigma Aldrich) in 95% ethanol. Coated slides were cured in a 110°C oven. cDNAs were applied to the coated glass substrate using a procedure described in USPN 5,807,522. One microliter of the cDNA at an average concentration of 100 ng/ μ l was loaded into the open capillary printing element by a high-speed robotic apparatus which then deposited about 5 nl of cDNA per slide.

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

X Preparation of Samples

LNCaP (ATCC, Manassus VA) is a prostate carcinoma cell line isolated from a lymph node biopsy of a 50-year-old male with metastatic prostate carcinoma. LNCaP cells express prostate specific antigens, produce prostatic acid phosphatase, and express androgen receptors. Gene expression profiles of LNCaP prostate carcinoma cells were compared to those of nontumorigenic primary prostate epithelial PrEC cells.

XI Expression of STEAPRP

WO 02/072596

PCT/US02/07593

Expression and purification of the protein are achieved using either a mammalian cell expression system or an insect cell expression system. The pUB6/V5-His vector system (Invitrogen, Carlsbad CA) is used to express STEAPRP in CHO cells. The vector contains the selectable bsd gene, multiple cloning sites, the promoter/enhancer sequence from the human ubiquitin C gene, a C-terminal V5 epitope for antibody detection with anti-V5 antibodies, and a C-terminal polyhistidine (6xHis) sequence for rapid purification on PROBOND resin (Invitrogen). Transformed cells are selected on media containing blasticidin.

Spodoptera frugiperda (SF9) insect cells are infected with recombinant *Autographica californica* nuclear polyhedrosis virus (baculovirus). The polyhedrin gene is replaced with the cDNA 10 by homologous recombination and the polyhedrin promoter drives cDNA transcription. The protein is synthesized as a fusion protein with 6xhis which enables purification as described above. Purified protein is used in the following activity and to make antibodies

XII Production of Antibodies

STEAPRP is purified using polyacrylamide gel electrophoresis and used to immunize mice or 15 rabbits. Antibodies are produced using the protocols below. Alternatively, the amino acid sequence of STEAPRP is analyzed using LASERGENE software (DNASTAR) to determine regions of high antigenicity. An antigenic epitope, usually found near the C-terminus or in a hydrophilic region is selected, synthesized, and used to raise antibodies. Typically, epitopes of about 15 residues in length are produced using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc-chemistry and 20 coupled to KLH (Sigma-Aldrich) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester to increase antigenicity.

Rabbits are immunized with the epitope-KLH complex in complete Freund's adjuvant. Immunizations are repeated at intervals thereafter in incomplete Freund's adjuvant. After a minimum of seven weeks for mouse or twelve weeks for rabbit, antisera are drawn and tested for antipeptide 25 activity. Testing involves binding the peptide to plastic, blocking with 1% bovine serum albumin, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG. Methods well known in the art are used to determine antibody titer and the amount of complex formation.

XIII Purification of Naturally Occurring Protein Using Specific Antibodies

30 Naturally occurring or recombinant protein is purified by immunoaffinity chromatography using antibodies which specifically bind the protein. An immunoaffinity column is constructed by covalently coupling the antibody to CNBr-activated SEPHAROSE resin (APB). Media containing the protein is passed over the immunoaffinity column, and the column is washed using high ionic strength buffers in the presence of detergent to allow preferential absorbance of the protein. After coupling, 35 the protein is eluted from the column using a buffer of pH 2-3 or a high concentration of urea or

WO 02/072596

PCT/US02/07053

thiocyanate ion to disrupt antibody/protein binding, and the protein is collected.

XIV Screening Molecules for Specific Binding with the cDNA or Protein

The cDNA, or fragments thereof, or the protein, or portions thereof, are labeled with ³²P-dCTP, Cy3-dCTP, or Cy5-dCTP (APB), or with BIODIPY or FITC (Molecular Probes, Eugene OR), respectively. Libraries of candidate molecules or compounds previously arranged on a substrate are incubated in the presence of labeled cDNA or protein. After incubation under conditions for either a nucleic acid or amino acid sequence, the substrate is washed, and any position on the substrate retaining label, which indicates specific binding or complex formation, is assayed, and the ligand is identified. Data obtained using different concentrations of the nucleic acid or protein are used to

calculate affinity between the labeled nucleic acid or protein and the bound molecule.

XV Two-Hybrid Screen

A yeast two-hybrid system, MATCHMAKER LexA Two-Hybrid system (Clontech Laboratories, Palo Alto CA), is used to screen for peptides that bind the protein of the invention. A cDNA encoding the protein is inserted into the multiple cloning site of a pLexA vector, ligated, and transformed into *E. coli*. cDNA, prepared from mRNA, is inserted into the multiple cloning site of a pB42AD vector, ligated, and transformed into *E. coli* to construct a cDNA library. The pLexA plasmid and pB42AD-cDNA library constructs are isolated from *E. coli* and used in a 2:1 ratio to co-transform competent yeast EGY48[p8op-lacZ] cells using a polyethylene glycol/lithium acetate protocol. Transformed yeast cells are plated on synthetic dropout (SD) media lacking histidine (-His), tryptophan (-Trp), and uracil (-Ura), and incubated at 30C until the colonies have grown up and are counted. The colonies are pooled in a minimal volume of 1x TE (pH 7.5), replated on SD/-His/-Leu/-Trp/-Ura media supplemented with 2% galactose (Gal), 1% raffinose (Raf), and 80 mg/ml 5-bromo-4-chloro-3-indolyl β-d-galactopyranoside (X-Gal), and subsequently examined for growth of blue colonies. Interaction between expressed protein and cDNA fusion proteins activates expression of a LEU2 reporter gene in EGY48 and produces colony growth on media lacking leucine (-Leu). Interaction also activates expression of β-galactosidase from the p8op-lacZ reporter construct that produces blue color in colonies grown on X-Gal.

Positive interactions between expressed protein and cDNA fusion proteins are verified by isolating individual positive colonies and growing them in SD/-Trp/-Ura liquid medium for 1 to 2 days at 30C. A sample of the culture is plated on SD/-Trp/-Ura media and incubated at 30C until colonies appear. The sample is replica-plated on SD/-Trp/-Ura and SD/-His/-Trp/-Ura plates. Colonies that grow on SD containing histidine but not on media lacking histidine have lost the pLexA plasmid. Histidine-requiring colonies are grown on SD/Gal/Raf/X-Gal/-Trp/-Ura, and white colonies are isolated and propagated. The pB42AD-cDNA plasmid, which contains a cDNA encoding a protein that physically interacts with the protein, is isolated from the yeast cells and characterized.

WO 02/072596

PCT/US02/07053

XVI STEAPRP Assay

The localization of STEAPRP in the prostate is detected by immunohistochemical analysis as described by Hubert *et al.* (*supra*). Prostate tissue sections (4-mm) are fixed with formalin and embedded in paraffin. Tissues are incubated with anti-STEAPRP antibodies, washed, and then treated
5 with biotinylated rabbit anti-sheep IgG. STEAPRP is visualized with avidin-conjugated horseradish peroxidase (Vector Laboratories, Burlingame CA).

All patents and publications mentioned in the specification are incorporated by reference herein. Various modifications and variations of the described method and system of the invention will
10 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 specific preferred 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 that are obvious to those skilled in the field of molecular biology or related fields are intended to be
15 within the scope of the following claims.

WO 02/072596

PCT/US02/07053

Tissue Category	Clone Count	Found in	Abs Abund	Pct Abund
Cardiovascular System	266190	0/68	0	0.0000
Connective Tissue	144645	0/47	0	0.0000
Digestive System	501101	3/148	3	0.0006
Embryonic Structures	106713	0/21	0	0.0000
Endocrine System	225386	2/53	2	0.0009
Exocrine Glands	254635	1/64	1	0.0004
Reproductive, Female	427284	2/106	2	0.0005
Reproductive, Male	448207	28/114	43	0.0096
Germ Cells	38282	0/5	0	0.0000
Hemic and Immune System	680277	2/159	3	0.0004
Liver	109378	1/35	2	0.0018
Musculoskeletal System	159280	2/47	3	0.0019
Nervous System	955753	9/198	12	0.0013
Pancreas	110207	1/24	2	0.0018
Respiratory System	390086	6/93	9	0.0023
Sense Organs	19256	0/8	0	0.0000
Skin	72292	0/15	0	0.0000
Stomatognathic System	12923	0/10	0	0.0000
Unclassified/Mixed	120926	1/13	1	0.0008
Urinary Tract	279062	2/64	2	0.0007
Totals	5321883	60/1292	85	0.0000

TABLE 1

WO 02/072596

PCT/US02/07053

Library ID	Clone Count	Library Description	Abs Abund	Pct Abund
PROGNOT19	3679	prostate, AH, mw/adenocA, M	4	0.1087
PROGNOT18	2201	prostate tumor, adenocA, 68M	2	0.0509
PROGNOT6	1157	prostate, PIN, mw/cancer, M	1	0.0864
PROGNOT26	3705	prostate, mw/adenocA, 65M	3	0.0810
PROGNOT1	3873	prostate, AH, mw/adenocA, 58M	3	0.0775
PROGNOT20	4550	prostate tumor, adenocA, 59M, SUB	3	0.0659
PROGNOT4	8553	prostate tumor, adenocA, 57M	3	0.0351
PROGNOT20	2995	prostate, AH, mw/adenocA, 65M	1	0.0334
PROGNOT21	3268	prostate tumor, adenocA, 61M	1	0.0306
PROGNOT519	4087	prostate tumor, adenocA, 59M, SUB	1	0.0245
PROGNOT12	7138	prostate tumor, adenocA, 65M	1	0.0140
PROGNOT6	8829	prostate, AH, mw/adenocA, 57M	1	0.0113

TABLE 2

WO 02/072596

PCT/US02/07053

mean log2 DE (Cys/Cys)	Cy3	Cy5
1.82	0	Human, REC Cells, Starved 24hr
3	0	Human, LNCaP Line, Starved 24hr, CA
1.86	0	Human, REC Cells, Starved 24hr
3.02	0	Human, LNCaP Line, Rich media 24hr, CA
2.5	12.45	Human, REC Cells
		Human, LNCaP Line, CA

TABLE 3

WO 02/072596

PCT/US02/07053

What is claimed is:

1. An isolated cDNA comprising a nucleic acid sequence encoding a protein having the amino acid sequence of SEQ ID NO:1, or the complement thereof.
2. An isolated cDNA comprising a nucleic acid sequence selected from:
 - 5 a) SEQ ID NO:2 or the complement thereof;
 - b) a fragment of SEQ ID NO:2 selected from SEQ ID NOs:3-9 or the complement thereof; and
 - c) a variant of SEQ ID NO:2 selected from SEQ ID NO:10 or the complement thereof.
3. A composition comprising the cDNA or the complement of the cDNA of claim 1 and a labeling moiety.
- 10 4. A vector comprising the cDNA of claim 1.
5. A host cell comprising the vector of claim 4.
6. A method for using a cDNA to produce a protein, the method comprising:
 - a) culturing the host cell of claim 5 under conditions for protein expression; and
 - b) recovering the protein from the host cell culture.
- 15 7. A method for using a cDNA to detect expression of a nucleic acid in a sample comprising:
 - a) hybridizing the composition of claim 3 to nucleic acids of the sample, thereby forming hybridization complexes; and
 - b) comparing hybridization complex formation with a standard, wherein the comparison indicates expression of the cDNA in the sample.
- 20 8. The method of claim 7 further comprising amplifying the nucleic acids of the sample prior to hybridization.
9. The method of claim 7 wherein the composition is attached to a substrate.
10. The method of claim 7 wherein the cDNA is differentially expressed when compared with a standard and is diagnostic of prostate hyperplasia or prostate cancer.
- 25 11. A method of using a cDNA to screen a plurality of molecules or compounds, the method comprising:
 - a) combining the cDNA of claim 1 with a plurality of molecules or compounds under conditions to allow specific binding; and
 - b) detecting specific binding, thereby identifying a molecule or compound which specifically
 - 30 binds the cDNA.
12. The method of claim 11 wherein the molecules or compounds are selected from DNA molecules, RNA molecules, peptide nucleic acids, artificial chromosome constructions, peptides, transcription factors, repressors, and regulatory molecules.
13. A purified protein or a portion thereof produced by the method of claim 6 and selected from:
 - 35 a) an amino acid sequence of SEQ ID NO:1;

WO 02/072596

PCT/US02/07053

- b) an antigenic epitope of SEQ ID NO:1; and
 - c) a biologically active portion of SEQ ID NO:1.
14. A composition comprising the protein of claim 13 and a pharmaceutical carrier.
15. A method for using a protein to screen a plurality of molecules or compounds to identify at least
5 one ligand, the method comprising:
- a) combining the protein of claim 13 with the molecules or compounds under conditions to allow specific binding; and
 - b) detecting specific binding, thereby identifying a ligand which specifically binds the protein.
16. The method of claim 15 wherein the molecules or compounds are selected from DNA molecules,
10 RNA molecules, peptide nucleic acids, peptides, proteins, mimetics, agonists, antagonists, antibodies, immunoglobulins, inhibitors, and drugs.
17. A method of using a protein to prepare and purify antibodies comprising:
- a) immunizing a animal with the protein of claim 15 under conditions to elicit an antibody response;
 - 15 b) isolating animal antibodies;
 - c) attaching the protein to a substrate;
 - d) contacting the substrate with isolated antibodies under conditions to allow specific binding to the protein;
 - e) dissociating the antibodies from the protein, thereby obtaining purified antibodies.
- 20 18. An antibody produced by the method of claim 17.
19. A method for using an antibody to diagnose conditions or diseases associated with expression of a protein, the method comprising:
- a) combining the antibody of claim 18 with a sample, thereby forming antibody:protein
complexes; and
 - 25 b) comparing complex formation with a standard, wherein the comparison indicates expression of the protein in the sample.
20. The method of claim 19 wherein expression is diagnostic of prostate hyperplasia or prostate cancer.
21. A method for preparing a monoclonal antibody with the specificity of the antibody of claim 18
30 comprising:
- a) immunizing a animal with a protein of SEQ ID NO:1 under conditions to elicit an antibody response;
 - b) isolating antibody-producing cells from the animal;
 - c) fusing the antibody-producing cells with immortalized cells in culture to form monoclonal
35 antibody producing hybridoma cells;

WO 02/072596

PCT/US02/07053

- d) culturing the hybridoma cells; and
 - e) isolating monoclonal antibodies from culture.
22. A monoclonal antibody produced by the method of claim 21.
23. A method for using an antibody to immunopurify a protein comprising:
- 5 a) attaching an antibody of claim 18 to a substrate,
- b) exposing the antibody to a sample containing protein under conditions to allow antibody:protein complexes to form,
- c) dissociating the protein from the complex, and
- d) collecting the purified protein.
- 10 24. A composition comprising an antibody of claim 18 and a labeling moiety.
25. A composition comprising an antibody of claim 18 and a pharmaceutical agent.

WO 02/072596

1/10

PCT/US02/07053

```

5'   9   18   27   36   45   54
    GG GGA AGC AGC TGG AGT GCG ACC GCC ACG GCA GCC ACC CTG CAA CCG CCA GTC

    63   72   81   90   99   108
    GGA GGT GCA GTC CGT AGG CCC TGG CCC CCG GGT GGG CCC TTG GGG AGT CCG CGC

    117  126  135  144  153  162
    CGC TCC CGA GGA GCT GCA AGG CTC GCC CCT GCC CGG CGT GGA GGG CGC GGG GGG

    171  180  189  198  207  216
    CGC GGA GAA AGT GAA GAG AGG AAA TTG GAA AAT TGT GAG TGG ACC TTC TCA TAC

    225  234  243  252  261  270
    TGC TCC TCC TTG CGT GGA AAA GGG GAA AGA ACT GCA TGC ATA TTA TTC AGC GTC

    279  288  297  306  315  324
    CTA TAT TCA AAG GAT ATT CTT GGT GAT CTT GGA AGT GTC CGT ATC ATG GAA TCA

    333  342  351  360  369  378
    ATC TCT ATG ATG GGA AGC CCT AAG AGC CTT AGT GAA ACT TGT TTA CCT AAT GGC

    I  S  M  M  G  S  P  K  S  L  S  E  T  C  L  P  N  G
    ---
    M  E  S

```

FIGURE 1A

WO 02/072596

2/10

PCT/US02/07053

387	ATA	AAT	GGT	ATC	AAA	GAT	GCA	AGG	AAG	GTC	ACT	GTA	GGT	GTG	ATT	GGA	AGT	GGA	432
	I	N	G	I	K	D	A	R	K	V	T	V	G	V	I	G	S	G	
441	GAT	TTT	GCC	AAA	TCC	TTG	ACC	ATT	CGA	CTT	ATT	AGA	TCC	GGC	TAT	CAT	GTG	GTC	486
	D	F	A	K	S	L	T	I	R	L	I	R	C	G	Y	H	V	V	
495	ATA	GGA	AGT	AGA	AAT	CCT	AAG	TTT	GCT	TCT	GAA	TTT	TTT	CCT	CAT	GTG	GTA	GAT	540
	I	G	S	R	N	P	K	F	A	S	E	F	F	P	H	V	V	D	
549	GTC	ACT	CAT	CAT	GAA	GAT	GCT	CTC	ACA	AAA	ACA	AAT	ATA	ATA	TTT	GTT	GCT	ATA	594
	V	T	H	H	E	D	A	L	T	K	T	N	I	I	F	V	A	I	
603	CAC	AGA	GAA	CAT	TAT	ACC	TCC	CTG	TGG	GAC	CTG	AGA	CAT	CTG	CTT	GTG	GGT	AAA	648
	H	R	E	H	Y	T	S	L	W	D	L	R	H	L	L	V	G	K	

FIGURE 1B

WO 02/072596

3/10

PCT/US02/07053

```

657      666      675      684      693      702
ATC CTG ATT GAT GTG AGC AAT AAC ATG AGG ATA AAC CAG TAC CCA GAA TCC AAT
-----
I  L  I  D  V  S  N  N  M  R  I  N  Q  Y  P  E  S  N
-----

711      720      729      738      747      756
GCT GAA TAT TTG GCT TCA TTA TTC CCA GAT TCT TTG ATT GTC AAA GGA TTT AAT
-----
A  E  Y  L  A  S  L  F  P  D  S  L  I  V  K  G  F  N
-----

765      774      783      792      801      810
GTT GTC TCA GCT TGG GCA CTT CAG TTA GCA CCT AAG GAT GCC AGC CGG CAG GTT
-----
V  V  S  A  W  A  L  Q  L  G  P  K  D  A  S  R  Q  V
-----

819      828      837      846      855      864
TAT ATA TGC AGC AAC AAT ATT CAA GCG CGA CAA CAG GTT ATT GAA CTT GCC CGC
-----
Y  I  C  S  N  N  I  Q  A  R  Q  Q  V  I  E  L  A  R
-----

873      882      891      900      909      918
CAG TTG AAT TTC ATT CCC ATT GAC TTG GGA TCC TTA TCA TCA GCC AGA GAG ATT
-----
Q  L  N  F  I  P  I  D  L  G  S  L  S  S  A  R  E  I

```

FIGURE 1C

WO 02/072596

4/10

PCT/US02/07053

927	GAA	ANT	TTA	CCC	CTA	CGA	CTC	TTT	ACT	CTC	TGG	AGA	GGG	CCA	GTG	GTG	GTA	GCT	972
	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
	E	N	L	P	L	R	L	F	T	L	W	R	G	P	V	V	V	A	
981	ATA	AGC	TTG	GCC	ACA	TTT	TTT	TTC	CTT	TAT	TCC	TTT	GTG	AGA	GAT	GTG	ATT	CAT	1026
	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
	I	S	L	A	T	F	F	F	L	Y	S	F	V	R	D	V	I	H	
1035	CCA	TAT	GCT	AGA	AAC	CAA	CAG	AGT	GAC	TTT	TAC	AAA	ATT	CCT	ATA	GAG	ATT	GTG	1080
	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
	P	Y	A	R	N	Q	Q	S	D	F	Y	K	I	P	I	E	I	V	
1089	AAT	AAA	ACC	TTA	CCT	ATA	GTT	GCC	ATT	ACT	TTG	CTC	TCC	CTA	GTA	TAC	CTC	GCA	1134
	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
	N	K	T	L	P	I	V	A	I	T	L	L	S	L	V	Y	L	A	
1143	GGT	CTT	CTG	GCA	GCT	GCT	TAT	CAA	CTT	TAT	TAC	GGC	ACC	AAG	TAT	AGG	AGA	TTT	1188
	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
	G	L	L	A	A	A	Y	Q	L	Y	Y	G	T	K	Y	R	R	F	

FIGURE 1D

WO 02/072596

5/10

PCT/US02/07053

1197	1206	1215	1224	1233	1242
CCA CCT TGG TTG GAA ACC TGG TTA CAG TGT AGA AAA CAG CTT GGA TTA CTA AGT					

P P W L E T W L Q C R K Q L G L L S					

1251	1260	1269	1278	1287	1296
TTT TTC TTC GCT ATG GTC CAT GTT GCC TAC AGC CTC TGC TTA CCG ATG AGA AGG					

F F F A M V H V A Y S L C L P M R R					

1305	1314	1323	1332	1341	1350
TCA GAG AGA TAT TTG TTT CTC AAC ATG GCT TAT CAG CAG GTT CAT GCA AAT ATT					

S E R Y L F L N M A Y Q Q V H A N I					

1359	1368	1377	1386	1395	1404
GAA AAC TCT TGG AAT GAG GAA GAA GTT TGG AGA ATT GAA ATG TAT ATC TCC TTT					

E N S W N E E E V W R I E M Y I S F					

1413	1422	1431	1440	1449	1458
GGC ATA ATG AGC CTT GGC TTA CTT TCC CTC CTG GCA GTC ACT TCT ATC CCT TCA					

G I M S L G L L S L L A V T S I P S					

FIGURE 1E

WO 02/072596

6/10

PCT/US02/07053

```

1467      1476      1485      1494      1503      1512
GTG AGC AAT GCT TTA AAC TGG AGA GAA TTC AGT TTT ATT CAG TCT ACA CTT GGA
-----
V S N A L N W R E F S F I Q S T L G
-----
1521      1530      1539      1548      1557      1566
TAT GTC GCT CTG CTC ATA AGT ACT TTC CAT GTT TTA ATT TAT GGA TGG AAA CGA
-----
Y V A L L I S T F H V L I Y G W K R
-----
1575      1584      1593      1602      1611      1620
GCT TTT GAG GAA GAG TAC TAC AGA TTT TAT ACA CCA CCA AAC TTT GTT CTT GCT
-----
A F E E Y Y R F Y T P P N F V L A
-----
1629      1638      1647      1656      1665      1674
CTT GTT TTG CCC TCA ATT GTA ATT CTG GGT AAG ATT ATT TTA TTC CTT CCA TGT
-----
L V L P S I V I L G K I I L F L P C
-----
1683      1692      1701      1710      1719      1728
ATA AGC CGA AAG CTA AAA CGA ATT AAA AAA GGC TGG GAA AAG AGC CAA TTT CTG
-----
I S R K K L K R I K K G W E K S Q F L
-----

```

FIGURE 1F

WO 02/072596

7/10

PCT/US02/07053

```
1737      1746      1755      1764      1773      1782
GAA GAA GGT ATT GGA GGA ACA ATT CCT CAT GTC TCC CCG GAG AGG GTC ACA GTA
---
E E G I G G T I P H V S P E R V T V
1791      1800      1809      1818      1827      1836
AAG TGA TGA TAA ATG GTG TTC ACA GCT GCC ATA TAA AGT TCT ACT CAT GCC ATT
---
M
1845      1854      1863      1872      1881      1890
ATT TTT ATG ACT TCT ACG TTC AGT TAC AAG TAT GCT GTC AAA TTA TCG TCG GTT
GA 3'
---
```

FIGURE 1G

WO 02/072596

8/10

PCT/US02/07053

1 MESISMMSGSPKSLSETCLPNGINGIKDA RK 7492448
 1 MES-----RK 96572948
 31 VTGVIGSGDFAKSLTIRLCGYHVVIGS 7492448
 6 ----- 96572948
 61 RNPKFASFPFPHVVDVTHHEDALTKTNIIFF 7492448
 6 -----DITNQEE----- 96572948
 91 VAIHREHYTSLWDLRHLLVGKILIDVSNM 7492448
 13 -----LWKMKP-----RNNL 96572948
 121 RINQYPESSNAEYLA SLFFPPDSLIVKGFNVVS 7492448
 23 EDDY LHKDTG-ETSM LKRPV L----- 96572948
 151 AWA LQ LGPKD ASRQVYICSNNIQARQVIE 7492448
 44 -----LHL-HQTAHADFFDCPSLQHTQE----- 96572948
 181 LARQLNFIPIIDLGLSSLSSAREIENLP LFT 7492448
 67 -----LFP 96572948
 211 LWRGPVVVAISLA TFFFLYSFVRDVIHPYA 7492448
 70 QWHLPIKIAAIIASLTFLYTLRREVIHPLA 96572948

FIGURE 2A

WO 02/072596

9/10

PCT/US02/07053

241 R N Q Q S D F Y K I P I E I V N K T L P I V A I T L L S L V 7492448
 100 T S H Q Q Y F Y K I P I L V I N K V L P M V S I T L L A L V 96572948

 271 Y L A G L L A A Y Q L Y Y G T K Y R R F P P W L E T W L Q 7492448
 130 Y L P G V I A A I V Q L H N G T K Y K K F P H W L D K W M L 96572948

 301 C R K Q L G L L S F F F A M V H V A Y S I C L P M R R S E R 7492448
 160 T R K Q F G L L S F F F A V L H A I Y S L S Y P M R R S Y R 96572948

 331 Y L F L N M A Y Q Q V H A N I E N S W N E E E V W R I E M Y 7492448
 190 Y K L L N W A Y Q Q V Q N K E D A W I E H D V W R M E I Y 96572948

 361 I S F G T M S L G L L S L L A V T S I P S V S N A L N W R E 7492448
 220 V S L G I V G L A I L A L L A V T S I P S V S D S L T W R E 96572948

 391 F S F I Q S T L G Y V A L L I S T F H V L I Y G W K R A F E 7492448
 250 F H Y I Q S K L G I V S L L L G T I H A L I F A W N K W I D 96572948

 421 E E Y Y R F Y T P P N F V L A L V L P S I V I L G K I I L F 7492448
 280 I K Q F V W Y T P P T F M I A V F L P I V V L I F K S I L F 96572948

FIGURE 2B

WO 02/072596

10/10

PCT/US02/07053

451	L	P	C	I	S	R	K	L	K	R	I	K	K	G	W	E	K	S	Q	F	L	E	E	G	I	G	T	I	P	7492448	
310	L	P	C	L	R	K	I	L	K	I	R	H	G	W	E	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	g6572948
481	H	V	S	P	E	R	V	-	T	V	M																				7492448
329	K	I	N	K	T	E	I	C	S	Q	L																				g6572948

FIGURE 2C

WO 02/072596

PCT/US02/07053

<110> INCYTE GENOMICS, INC.
 LAL, Frestia G.
 FARIS, Mary
 CHEN, Hui-Mei
 ISON, Craig H.

<120> STEAP-RELATED PROTEIN

<130> PC-0037 PCT

<140> To Be Assigned

<141> Herewith

<151> 09/802,520

<152> 2001-03-09

<160> 11

<170> PERL Program

<210> 1

<211> 490

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7492448CD1

<400> 1

```

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

```

WO 02/072596

PCT/US02/07053

```

245      250      255
Asn Lys Thr Leu Pro Ile Val Ala Ile Thr Leu Leu Ser Leu Val
260      265      270
Tyr Leu Ala Gly Leu Leu Ala Ala Ala Tyr Gln Leu Tyr Tyr Gly
275      280      285
Thr Lys Tyr Arg Arg Phe Pro Pro Trp Leu Glu Thr Trp Leu Gln
290      295      300
Cys Arg Lys Gln Leu Gly Leu Leu Ser Phe Phe Phe Ala Met Val
305      310      315
His Val Ala Tyr Ser Leu Cys Leu Pro Met Arg Arg Ser Glu Arg
320      325      330
Tyr Leu Phe Leu Asn Met Ala Tyr Gln Gln Val His Ala Asn Ile
335      340      345
Glu Asn Ser Trp Asn Glu Glu Glu Val Trp Arg Ile Glu Met Tyr
350      355      360
Ile Ser Phe Gly Ile Met Ser Leu Gly Leu Leu Ser Leu Leu Ala
365      370      375
Val Thr Ser Ile Pro Ser Val Ser Asn Ala Leu Asn Trp Arg Glu
380      385      390
Phe Ser Phe Ile Gln Ser Thr Leu Gly Tyr Val Ala Leu Leu Ile
395      400      405
Ser Thr Phe His Val Leu Ile Tyr Gly Trp Lys Arg Ala Phe Glu
410      415      420
Glu Glu Tyr Tyr Arg Phe Tyr Thr Pro Pro Asn Phe Val Leu Ala
425      430      435
Leu Val Leu Pro Ser Ile Val Ile Leu Gly Lys Ile Ile Leu Phe
440      445      450
Leu Pro Cys Ile Ser Arg Lys Leu Lys Arg Ile Lys Lys Gly Trp
455      460      465
Glu Lys Ser Gln Phe Leu Glu Glu Gly Ile Gly Gly Thr Ile Pro
470      475      480
His Val Ser Pro Glu Arg Val Thr Val Met
485      490

```

<210> 2

<211> 1891

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Inocyte ID No: 7492448CB1

<400> 2

```

ggggaagcag ctggagtgcg accgccacgg cagccaccot gcaaccgcca gtggaggggtg 60
cagtccgtag gccctggccc cggggtgggc ccttggggag tggcgccgcg tccgagggag 120
ctgcaaggct cgcctctgcc cggcgtggag ggcgcggggg gcgcggagaa agtgaagaga 180
ggaaattgga aaattgtgag tggaccttct gatactgctc ctcttgcgtt ggaaaagggg 240
aaagaactgc atgcatatta ttcagcgttc tatattcaaa ggatattctt ggtgatcttg 300
gaagtgtccg tatcatggaa tcaatctcta tgatgggaag ccctaagagc cttagtga 360
cttgtttacc taatggcata aatgggtatca aagatgcaag gaaggtcact gtagggtgta 420
ttggaagtgg agattttgcc aaatccttga ccattcgact tattagatgc ggctatcatg 480
tggtcacagg aagttagaat cctaagtttg ctcttgaatt ttctctcat gtgtagatg 540
tcctctgata tgaagatgct ctcaaaaaa caatataat attgttgct atacacagag 600
aacattatac ctccctgtgg gaactgagac atctgcttgt gggtaaaac ctgattgatg 660
tgagcaataa catgaggata aaccagtcac cagaatccaa tgcgtaatat ttggtctcat 720
tattcccaga ttctttgatt gtcaaaggat ttaatgttgt ctacgcttgg gcacttccagt 780
taggaacctaa gtagtcacgc cggcaggttt atatatgcag caacaatatt caagtcacag 840
aacaggttat tgaacttgcc cggcagttga atttcattcc cattgaactg ggaaccttat 900
catcagccag agagattgaa aatttaccoc taagactctt tactctctgg agagggccag 960
tggtggtagc tataagcttg gccacatttt ttttccttta ttcccttgcc agagatgtga 1020
ttcatccata tgctagaac caacagatg acttttcaaa aattctcata gagattgtga

```

WO 02/072596

PCT/US02/07053

```

1080
ataaaacott aacctatggt gccattactt tgcctctccct agtataccct gcagggtcttc
1140
tggcagctgc ttatcaactt tattacggca ccaagtatag gagatttcca ccttggttgg
1200
aaactctggtt acagtgtaga aaacagcttg gattactaag ttttttcttc gctatggtcc
1260
atgttgccca cagcctctgc ttaccgatga gaaggtcaga gagatatttg tttctcaaca
1320
tggcttatca gcagggttcct gcaaatattg aaaactcttg gaatgaggaa gaagtctgga
1380
gaattgaaat gtatatctcc tttggcataa tgagccttgg ctactttcc ctcctggcag
1440
tcacttttat ccttcagtg agcaatgctt taaactggag agaattcagt tttattcagt
1500
ctacacttgg atatgtgct ctgctcataa gtactttcca tgttttaatt tatggatgga
1560
aacagcttt tgaggagag tactacagat tttatacaco accaaacttt gttcttgctc
1620
ttgttttggc ctcaattgta attctgggta agattatatt attccttcca tgtataagcc
1680
gaaagctaaa acgaattaaa aaaggctggg aaagagcca atttctggaa gaaggatttg
1740
gaggaaacaat tctcatgtc tcccggaga gggctacagt aatgtgatga taaatggtgt
1800
tcacagctgc catataaagt tctactcatg ccattatatt tatgaactct acgttcagtt
1860
acaagtatgc tgtcaatta tcgtgggttg a
1691

<210> 3
<211> 517
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7100809H1

<400> 3
ggggaagcag ctggagtgcg accgccacgg cagccacect gcaaccgcca gtcggaggtg 60
cagtcogtag gccctggccc ccgggtgggc ccttgggaag tgggcgcgcg tcccgaggag 120
ctgcaaggct cgcctctgcc cggcgtggag ggcggggagg gcgcggagaa agtgaagaga 180
ggaaattgga aaattgtgag tggaccttct gatactgctc ctcttgcgt ggaaagggg 240
aaagaactcg atcatatta ttcagctctc tatattcaaa ggatattctt ggtgatcttg 300
gaagtgtccg tatcatggaa tcaatctota tgatgggaag ccctaagagc cttagtga 360
cttgtttacc taatggcata aatggtatca aagatgcaag gaaggtcact gtaggtgtga 420
ttggaagtgg agattttgcc aaatccttga ccattcgact tattagatgc ggctatcatg 480
tggtcatagg aagtagaaat cctaagttgg cttctga 517

<210> 4
<211> 493
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6912820J1

<400> 4
ggtcactgta ggtgtgattg gaagtggaga ttttgccaaa tccttgacca ttogacttat 60
tagatgcggc tatcatgtgg tcatagggag tagaaatcct aagtttgctt ctgaattttt 120
tctcatgttg tagatgtca ctcatcatga agatgctctc acaaaaacaa atataatatt 180

```

WO 02/072596

PCT/US02/07053

```

tgttgctata cacagagAAC attatacctc cctgtgggac ctgagacatc tgcttgtggg 240
taaaatcccg attgatgtga gcaataacat gaggataaac cagtaccag aatccaatgc 300
tgaatatttg gcttcattat tcccagatcc ttgtattgtc aaaggattta atgttgcttc 360
agcttgga cttcagttag gacctaaagg tgccagccgg caggtttata tatgcagcaa 420
caatattcaa ggcggacaac aggttatiga acttgccgcg cagttgaatt tcaatccat 480
tgacttggga tcc

```

```

<210> 5
<211> 403
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 4647117F6

```

```

<220>
<221> unsure
<222> 316, 321, 339
<223> a, t, c, g, or other

```

```

<400> 5
cccagattct ttgattgtca aaggatttaa tgtgtctcga gcttggggac ttcagttagg 60
accctaaggat gccagccggc aggtttatat atgcagcaac aatattcaag cgcgacaaca 120
ggttattgaa cttgcccgcc agttgaattt cattcccatc gacttgggat ccttatcatc 180
agccagagag attgaaaatt taccocctacg actctttact ctctggagag ggccagtggg 240
ggtagctata agcttggcca catTTTTTTT cctttattcc ttgtcagag atgtgattca 300
tcataatgct agaaancaac ngagtgaact ttacaacnt tcatagaga ttgtgaataa 360
aaccttacct atagtgtcca ttactttgct cccctagta tac

```

```

<210> 6
<211> 560
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7004364H1

```

```

<400> 6
acattttttt tctttgatgc ctttgtcaga gatgtgatcc atccatatgc tagaaaccaa 60
cagagtgaact ttcaacaat tcctatagag attgtgaata aaacctcacc tatagttgct 120
attactttgc tctccctagt atacctgcga ggtcttctgg cagctgctta tcaactttat 180
tacggcaaca agtataggag atttccacct tggttggaaa cctggttaca gtgtagaana 240
cagcttggat tactaagttt tatcttgcgt atggtccatg ttgcctacag cctctgttta 300
ccgatgagaa ggtcagagag atatttgttt ctcaacatgg ctatcagca ggttcacgca 360
aatattgaaa actcttggaa tgaggaagaa gtttggagaa ttgaaatgta tatctccttt 420
ggcataatga gccttggctt actttccctc ctggcagta cttctatccc ttcagtgaac 480
aatgctttta actggagaga attcagtttt attcagttta cacttggata tgcctctctg 540
ctcataagta ctttccatgt

```

```

<210> 7
<211> 265
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 70351677D1

```

```

<400> 7
ctcagtcctgg gtatctgcaa actgcaaaag atccagaatt acaattgagg gcaaaacaag 60
agcaagaaca aagtttgggt gtgtataaaa tctgtagtao tcttctctca aagctcgttt 120

```

WO 02/072596

PCT/US02/07053

```
ccatccataa attaaaacat ggaagactat tatgagcaga ggcacatata caagtgtaga 180
cgaataaaaa ctgaattctc tccagtttaa agcattgctc actgaaggga tagaagtgac 240
tgccaggagg gaaagtaagc caagg                                     265
```

<210> 8
<211> 204
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4108079H1

<220>
<221> unsure
<222> 45, 83, 132
<223> a, t, c, g, or other

```
<400> 8
cagagtttat acaccaccaa actttgttct tgetcgtgtt ttgencctcg gtgtaattct 60
ggggaagatt gttttattcc ttingtgata agggcgaagc taaaacgaat taagaaagcc 120
tggggaaaga gncogatttc tggagaaggg tctgggaggg acaattcgca tgtcgcccg 180
gagaggggtca cagtaatggg atga                                     204
```

<210> 9
<211> 265
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4659848H1

```
<400> 9
ccggagaggg tcacagtaat gtgatgataa atgggtgttca cagctgccat ataaagttct 60
actaatgcca ttatttttat gactttctacg ttcagttaca agtatgctgt caaattratcg 120
tggtttgaaa ctgtttaaag gagatttcaa ctgacttagt gatagagttt tottcaagtt 180
aattttcaca aatgtcatgt ttgccaatat gaatttttct agtcaacata ttattgtaat 240
ttaggtatgt tttgttttgt ttgc                                     265
```

<210> 10
<211> 525
<212> DNA
<213> Rattus norvegicus

<220>
<221> misc_feature
<223> Incyte ID No: 702819778T1

```
<400> 10
gggatgtgta atgttctcta tggatagcca cgaatattat atttgtcttc gttaaagcgt 60
cttcatgggtg ggtgacgtct accaatgag gaaaaaacctc agacgcgaac ttaggatttc 120
tgctttcgat gaccacgtga tagccgcacc tgataagccg aatggtcaga gacttggcaa 180
aatccccact tctatcaccc cccacgggtg acttctctgc gcttttgata cgttttatgc 240
catatagccaa aaacgtctcc aggtctttag ggcttcccat catagagatg gattccatgg 300
tagagactct tctaagatca ccaggaatgc cctgggaact ttaaggtgta gctttctact 360
cagaggaagct ggaggagggc tcttctggcg ctgctggact ctggaaactgc ctacgtgtag 420
tgaggagggc ctccgcgcgc tctctctccg gccacggctg cagcgcgcgc cegtggctcc 480
ctcgccccaa gggccgcgcg agctcccggg cctacggagt gctcc                                     525
```

<210> 11
<211> 339
<212> PRT

WO 02/072596

PCT/US02/07053

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: g6572948

<400> 11

```

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

```

```

Val Leu Ile Phe Lys Ser Ile Leu Phe Leu Pro Cys Leu Arg Lys
305    310    315
Lys Ile Leu Lys Ile Arg His Gly Trp Glu Asp Val Thr Lys Ile
320    325    330
Asn Lys Thr Glu Ile Cys Ser Gln Leu
335

```


【国際調査報告】

INTERNATIONAL SEARCH REPORT		International application No. PCT/US02/07053
A. CLASSIFICATION OF SUBJECT MATTER IPC(7) : C07H 21/02, 21/04; C12P 21/06; C12N 5/00 US CL : 536/23.1, 23.5; 435/69.1, 325 According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) U.S. : 536/23.1, 23.5; 435/69.1, 325 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) MPRCH		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6,329,503 B1 (AFAR et al) 11 December 2001, SEQ ID NO:8.	1
P		3-6
X	Database Genbank, Accession No. AC002064, GATTUNG et al. Gene Sequence. 09 May 1997. See entire document.	2
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent 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 "A" document member of the same patent family		
Date of the actual completion of the international search 16 June 2002 (16.06.2002)		Date of mailing of the international search report 22 AUG 2002
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230		Authorized officer ANTHONY CAPUTO Telephone No. 703-308-6246

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

INTERNATIONAL SEARCH REPORT		International application No. PCT/US02/07053
Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)		
This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:		
1.	<input type="checkbox"/>	Claim Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
2.	<input type="checkbox"/>	Claim Nos.: because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3.	<input type="checkbox"/>	Claim Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)		
This International Searching Authority found multiple inventions in this international application, as follows: Please See Continuation Sheet		
1.	<input type="checkbox"/>	As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2.	<input type="checkbox"/>	As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3.	<input type="checkbox"/>	As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4.	<input checked="" type="checkbox"/>	No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-6
Remark on Protest		
	<input type="checkbox"/>	The additional search fees were accompanied by the applicant's protest.
	<input type="checkbox"/>	No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/07053

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I. Claims 1-6, drawn to a nucleic acid sequence of SEQ ID NO:2, or the complement thereof, or a nucleic acid sequence encoding SEQ ID NO:1, fragments selected from SEQ ID NO:3-9, and a variant of SEQ ID NO:10, a vector comprising a nucleic acid encoding SEQ ID NO:1, a host cell comprising said vector, and a method for producing a protein, classified in class 536, subclass 23.1.

Group II. Claims 7, 9, drawn to a method for detecting the expression of a nucleic acid, using hybridization, classified in class 435, subclass 6.

Group III. Claim 8, drawn to a method for detecting the expression of a nucleic acid, using amplification, classified in class 435, subclass 6.

Group IV. Claim 10, drawn to a method for detecting prostate hyperplasia or prostate cancer, classified in class 435, subclass 6.

Group V. Claims 11-12, drawn to a method for screening compounds that bind specifically to a nucleic acid encoding SEQ ID NO:1, classified in class 435, subclass 6.

Group VI. Claims 13-14, drawn to a protein of SEQ ID NO:1 or fragments thereof, classified in class 530, subclass 350.

Group VII. Claims 15-16, drawn to a method for screening compounds that specifically binds to SEQ ID NO:1, classified in class 435, subclass 7.1.

Group VIII. Claims 18,21-24, drawn to an antibody, classified in class 530, subclass 387.1.

Group IX. Claims 19-20, drawn to a method for detecting prostate hyperplasia or prostate cancer, using an antibody, classified in class 435, subclass 7.1.

In addition, upon election of group I, further election of the following species is required:

Full length sequence of SEQ ID NO:2, or fragments or a variant thereof

Upon election of fragments of SEQ ID NO:2, further election of the following species is required:

SEQ ID Nos: 3-9.

The inventions listed as Groups I-IX do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

According to PCT Rule 13.2, unity of invention exists only when the shared same or corresponding technical feature is a contribution over the prior art. The inventions listed as groups I-IX do not relate to a single general inventive concept because they lack the same or corresponding special technical feature. The technical feature of group I, a complement of SEQ ID NO:2, which is shown to be the same as the inherent complement of a nucleic acid molecule of SEQ ID NO:8 taught by PN=6329503, lacks novelty and does not make a contribution over the prior art.

The species do not share the same technical feature, because they have different structure.

フロントページの続き

(51)Int.Cl. ⁷	F I	テーマコード(参考)
A 6 1 P 13/08	A 6 1 P 13/08	4 C 0 8 4
A 6 1 P 35/00	A 6 1 P 35/00	4 C 0 8 5
C 0 7 K 14/705	C 0 7 K 14/705	4 C 0 8 6
C 0 7 K 16/28	C 0 7 K 16/28	4 H 0 4 5
C 1 2 N 1/15	C 1 2 N 1/15	
C 1 2 N 1/19	C 1 2 N 1/19	
C 1 2 N 1/21	C 1 2 N 1/21	
C 1 2 N 5/10	C 1 2 P 21/02	C
C 1 2 P 21/02	C 1 2 P 21/08	
C 1 2 P 21/08	C 1 2 Q 1/68	A
C 1 2 Q 1/68	G 0 1 N 33/15	Z
G 0 1 N 33/15	G 0 1 N 33/50	Z
G 0 1 N 33/50	G 0 1 N 33/53	D
G 0 1 N 33/53	G 0 1 N 33/566	
G 0 1 N 33/566	A 6 1 K 37/02	
	C 1 2 N 5/00	A

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT, BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN, TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EC,EE,ES,FI,GB,GD,GE, GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NO,NZ,OM,PH,P L,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZM,ZW

(72)発明者 ファリス、メアリー

アメリカ合衆国カリフォルニア州 9 0 0 7 7 ・ ロスアンゼルス・アルマデンコート 2 5 3 8

(72)発明者 チェン、ヒュエイ - メイ

アメリカ合衆国カリフォルニア州 9 5 0 5 6 ・ サンタクララ・ライルコート 2 4 2 5

(72)発明者 アイソン、クレイグ・エイチ

アメリカ合衆国カリフォルニア州 9 5 1 1 8 ・ サンノゼ・ウェザーズフィールドウェイ 1 2 4 2

F ターム(参考) 2G045 AA34 AA35 BB14 BB20 BB41 BB48 CB01 DA13 DA36 FA12

FA16 FA29 FB02 FB03 FB12 GC15

4B024 AA01 AA12 BA31 BA44 CA01 CA04 GA11 HA12

4B063 QA19 QQ02 QQ08 QQ53 QQ79 QQ91 QR08 QR32 QR42 QR48

QR55 QR62 QS25 QS33 QS34 QX02

4B064 AG27 AG31 CA19 CC24 CE13 DA01 DA14

4B065 AA93Y AB01 AC14 BA02 CA24 CA25 CA44 CA46

4C084 AA02 AA07 AA13 BA01 BA22 BA23 CA53 CA59 NA14 ZA81

ZB26

4C085 AA13 AA14 BB31 DD62 EE01

4C086 AA01 EA16 MA01 MA04 NA14 ZA81 ZB26

4H045 AA10 AA11 BA10 CA44 DA50 DA76 EA28 EA51 FA72 FA74

专利名称(译)	STEAP相关蛋白		
公开(公告)号	JP2004526444A	公开(公告)日	2004-09-02
申请号	JP2002571510	申请日	2002-03-07
[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	ラルプリーティジー ファリスメアリー チェンヒュエイメイ アイソンクレイグエイチ		
发明人	ラル、プリーティ・ジー ファリス、メアリー チェン、ヒュエイ-メイ アイソン、クレイグ・エイチ		
IPC分类号	G01N33/50 A61K31/7088 A61K38/00 A61K39/395 A61K48/00 A61P13/08 A61P35/00 C07K14/705 C07K16/28 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N15/09 C12P21/02 C12P21/08 C12Q1/68 G01N33/15 G01N33/53 G01N33/566		
CPC分类号	A61P13/08 C07K14/705		
FI分类号	C12N15/00.ZNA.A A61K31/7088 A61K39/395.D A61K39/395.N A61K48/00 A61P13/08 A61P35/00 C07K14/705 C07K16/28 C12N1/15 C12N1/19 C12N1/21 C12P21/02.C C12P21/08 C12Q1/68.A G01N33/15.Z G01N33/50.Z G01N33/53.D G01N33/566 A61K37/02 C12N5/00.A		
F-TERM分类号	2G045/AA34 2G045/AA35 2G045/BB14 2G045/BB20 2G045/BB41 2G045/BB48 2G045/CB01 2G045 /DA13 2G045/DA36 2G045/FA12 2G045/FA16 2G045/FA29 2G045/FB02 2G045/FB03 2G045/FB12 2G045/GC15 4B024/AA01 4B024/AA12 4B024/BA31 4B024/BA44 4B024/CA01 4B024/CA04 4B024 /GA11 4B024/HA12 4B063/QA19 4B063/QQ02 4B063/QQ08 4B063/QQ53 4B063/QQ79 4B063/QQ91 4B063/QR08 4B063/QR32 4B063/QR42 4B063/QR48 4B063/QR55 4B063/QR62 4B063/QS25 4B063 /QS33 4B063/QS34 4B063/QX02 4B064/AG27 4B064/AG31 4B064/CA19 4B064/CC24 4B064/CE13 4B064/DA01 4B064/DA14 4B065/AA93Y 4B065/AB01 4B065/AC14 4B065/BA02 4B065/CA24 4B065 /CA25 4B065/CA44 4B065/CA46 4C084/AA02 4C084/AA07 4C084/AA13 4C084/BA01 4C084/BA22 4C084/BA23 4C084/CA53 4C084/CA59 4C084/NA14 4C084/ZA81 4C084/ZB26 4C085/AA13 4C085 /AA14 4C085/BB31 4C085/DD62 4C085/EE01 4C086/AA01 4C086/EA16 4C086/MA01 4C086/MA04 4C086/NA14 4C086/ZA81 4C086/ZB26 4H045/AA10 4H045/AA11 4H045/BA10 4H045/CA44 4H045 /DA50 4H045/DA76 4H045/EA28 4H045/EA51 4H045/FA72 4H045/FA74		
优先权	09/802520 2001-03-09 US		
外部链接	Espacenet		

摘要(译)

本发明提供了编码STEAP相关蛋白的cDNA。它还提供了cDNA，其片段，补体和变体以及编码的蛋白质，其部分和其抗体在诊断和治疗前列腺细胞增生性疾病，特别是前列腺增生和前列腺癌中的用途。本发明还提供了用于产生蛋白质的表达载体和宿主细胞以及转基因模型系统。

(51) Int. Cl. ⁷	F I	テーマコード (参考)
C 1 2 N 15/09	C 1 2 N 15/00 Z N A A	2 G 0 4 5
A 6 1 K 31/7088	A 6 1 K 31/7088	4 B 0 2 4
A 6 1 K 38/00	A 6 1 K 39/385 D	4 B 0 6 3
A 6 1 K 39/395	A 6 1 K 39/395 N	4 B 0 6 4
A 6 1 K 48/00	A 6 1 K 48/00	4 B 0 6 5
審査請求 未請求 予備審査請求 未請求 (全 100 頁) 最終頁に付		

(21) 出願番号	特願2002-571510 (P2002-571510)	(71) 出願人	301005050
(86) (22) 出願日	平成14年3月7日 (2002.3.7)		インサイト・ゲノミックス・インコーポ
(85) 翻訳文提出日	平成15年9月8日 (2003.9.8)		イテッド
(86) 国際出願番号	PCT/US2002/007053		アメリカ合衆国カリフォルニア州943
(87) 国際公開番号	W02002/072596		4・パロアルト・ボートードライブ 3
(87) 国際公開日	平成14年9月19日 (2002.9.19)		60
(31) 優先権主張番号	09/802,520	(74) 代理人	100089266
(32) 優先日	平成13年3月9日 (2001.3.9)		弁理士 大島 陽一
(33) 優先権主張国	米国 (US)	(72) 発明者	ラル、ブリーティ・ジー
			アメリカ合衆国カリフォルニア州95C
			6・サンタクララ・ビーオーボックス
			142

最終頁に続く