

(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2004-507242

(P2004-507242A)

(43) 公表日 平成16年3月11日(2004.3.11)

(51) Int.Cl. ⁷	F I	テーマコード (参考)
C 1 2 N 15/09	C 1 2 N 15/00	Z N A A 2 G O 4 5
A 6 1 K 38/00	A 6 1 K 39/395	N 4 B O 2 4
A 6 1 K 39/395	A 6 1 K 39/395	T 4 B O 6 3
C O 7 K 14/82	C O 7 K 14/82	4 B O 6 4
C O 7 K 16/32	C O 7 K 16/32	4 B O 6 5
	審査請求 未請求 予備審査請求 有 (全 106 頁) 最終頁に続く	

(21) 出願番号 特願2002-522268 (P2002-522268)
 (86) (22) 出願日 平成13年8月23日 (2001.8.23)
 (85) 翻訳文提出日 平成15年2月17日 (2003.2.17)
 (86) 国際出願番号 PCT/US2001/026317
 (87) 国際公開番号 W02002/016595
 (87) 国際公開日 平成14年2月28日 (2002.2.28)
 (31) 優先権主張番号 09/645, 961
 (32) 優先日 平成12年8月24日 (2000.8.24)
 (33) 優先権主張国 米国 (US)

(71) 出願人 301005050
 インサイト・ゲノミックス・インコーポレ
 イテッド
 アメリカ合衆国カリフォルニア州9430
 4・パロアルト・ポータードライブ 31
 60
 (74) 代理人 100089266
 弁理士 大島 陽一
 (72) 発明者 ウォーカー、マイケル・ジー
 アメリカ合衆国カリフォルニア州9408
 9・サニーベイル・#80・ボレガスアベ
 ニュー 1050

最終頁に続く

(54) 【発明の名称】 腎臓特異タンパク質

(57) 【要約】

本発明は、腎臓特異タンパク質をコード化する哺乳動物
 cDNAを提供する。また腎臓の疾患の診断、治療のた
 めのcDNA、断片、相補体、その変異体およびコード
 化されたタンパク質またはその部分、その抗体の使用を
 も提供する。

```

5' GTGC TCT CTT CCA AGG CTG TAG GAG TTC TGG AGC TGC TGG CTG GAG AGG AGG GTG
10 19 28 37 46 55
64 73 82 91 100 109
GAC GAA GCT CTC TCT ATG AAG ACA TCC TGA GAG GAC TTG TCA GGC CTG AAT ATG
118 127 136 145 154 163
CAC TGG CTG CGA AAA CTT CAG CGA CTT TGC ACC CTG TGG GGT ACT CAG ATG TCC
H W L R K V Q G L C T L W G Y Q M S
172 181 190 199 208 217
AGC CGC ACT CTC TAC ATT AAT AGT AGG CAA CTG CTG TCC CTG CAG TGG GGC CAC
S R T L Y I N S R Q L V S L Q W G H
226 235 244 253 262 271
CA? GAA CTC CTG CCC AAG TTT AAC TTT GCT AGT GAT GTG TTG GAT CAC TGG GCT
Q E V P A K F N F A S D V L D H W A
280 289 298 307 316 325
GAC ATG GAG AAG GAT GAC AAG TGA CTC CCA AGC CCA GCC CTG TGG TGG GTG AAT
D M E K A G K R L P S P A L W W V N
334 343 352 361 370 379
GGG AAG GGG AAG GAA TTA ATG TGG AAT TTC AGA GAA CTG AGT GAA AAC AGC CAG
G K G K E L M W N F R E L S E N S Q

```

【特許請求の範囲】

【請求項 1】

SEQ ID NO : 1 のアミノ酸配列あるいは cDNA の相補体を有するタンパク質をコードする核酸配列から成る単離した cDNA

【請求項 2】

核酸配列から成る単離した cDNA であり、下記の a) 乃至 c) から選択される。

a) SEQ ID NO : 2 及びその相補体

b) SEQ ID NO : 2 の断片またはその相補体

c) SEQ ID NO : 1 に 83 % の配列同一性を有する SEQ ID NO : 2 の変異体、またはその相補体。

10

【請求項 3】

請求項 1 に記載の cDNA 及び標識成分を含む化合物。

【請求項 4】

請求項 1 に記載の cDNA を含むベクター。

【請求項 5】

請求項 4 に記載のベクターを含む宿主細胞。

【請求項 6】

タンパク質を産生する cDNA の使用に関する方法であり、

a) タンパク質発現のための条件下で請求項 5 に記載の宿主細胞を培養する方法

b) 宿主細胞株からタンパク質を回収する方法を含む。

20

【請求項 7】

サンプル中の核酸の発現を検出する cDNA を使用する方法であり、

a) 少なくとも 1 つのハイブリダイゼーション複合体を形成する条件下で、請求項 3 に記載の組成物をサンプル中の核酸に対してハイブリダイズさせる過程を含み、及び

b) ハイブリダイゼーション複合体形成を検出する過程を含み、複合体形成はサンプル中の cDNA の発現を示すことを特徴とする。

【請求項 8】

ハイブリダイゼーション前にサンプル中の核酸の増幅する過程をも含む請求項 7 に記載の方法。

【請求項 9】

組成物が基質に付着することを特徴とする請求項 7 に記載の方法。

30

【請求項 10】

複合体形成を少なくとも 1 つの標準と比較して、示差発現を決定することを特徴とする、請求項 7 に記載の方法。

【請求項 11】

cDNA を用いて複数の分子か化合物をスクリーニングする方法であり、

a) 特異結合を可能とする条件下で、請求項 1 に記載の cDNA と複数の分子あるいは化合物を結合する過程を含み、

b) 特異結合の検出および、それにより cDNA を特異結合する分子あるいは化合物を同定する過程を含む。

40

【請求項 12】

請求項 11 に記載の方法であり、分子または化合物を DNA 分子、RNA 分子、ペプチド核酸、人工形成の染色体、ペプチド、転写因子、リプレッサー、調節分子から選択することを特徴とする。

【請求項 13】

請求項 6 に記載の方法で生成された精製されたタンパク質あるいはその部分であり、下記の a) - c) から選択される。

a) SEQ ID NO : 1 のアミノ酸配列、

b) SEQ ID NO : 1 の抗原性エピトープ。

【請求項 14】

50

請求項 13 に記載のタンパク質及び医薬用担体を含む組成物。

【請求項 15】

複数の分子あるいは化合物をスクリーニングするタンパク質を用いて、少なくとも 1 つのリガンドを同定する方法であり、下記の a) - b) を含む。

a) 特異結合を可能とする条件下で、請求項 13 に記載のタンパク質と複数の分子あるいは化合物を結合する過程を含み、

b) 特異結合の検出および、それによりタンパク質を特異結合するリガンドを同定する過程を含む。

【請求項 16】

請求項 15 に記載の方法であり、分子または化合物を 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、薬剤から選択することを特徴とする。

【請求項 17】

タンパク質を用いて抗体を調製及び精製する方法であり、

a) 抗体反応を誘発する条件下で、請求項 15 に記載のタンパク質で動物を免疫化する過程、

b) 動物抗体を単離する過程、

c) 前記タンパク質を基質に付着する過程、

d) タンパク質への特異結合を可能にする条件下で、基質と単離した抗体とを接触させる過程、

e) 抗体をタンパク質から解離する過程を含み、それによって精製した抗体を得る。

【請求項 18】

請求項 17 に記載の方法で産出した抗体。

【請求項 19】

タンパク質の発現と関連する症状または疾患を診断するために抗体を用いる方法であり、

a) 請求項 18 に記載の抗体をサンプルと結合する過程を含む。それにより抗体：タンパク質化合物を形成する。

b) サンプル中のタンパク質の発現を示す比較を特徴とする、化合物形成を標準と比較する過程をも含む。

【請求項 20】

腎細胞癌の診断に寄与する発現を特徴とする請求項 19 に記載の方法。

【請求項 21】

請求項 18 に記載の抗体及び医薬用担体を含む医薬品組成物。

【請求項 22】

腎細胞癌の治療が必要な患者に請求項 21 の組成物を有効量投与することを含むことを特徴とする腎細胞癌の治療方法。

【発明の詳細な説明】

【0001】

(技術分野)

本発明は、腎細胞癌 (KSRCC) の診断に用いる腎臓特異タンパク質をコード化する哺乳類 cDNA および腎臓の疾患の診断と治療における cDNA とコード化したタンパク質の使用に関連する。

【0002】

(発明の背景)

生物間の系統発生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。

【0003】

ヒトの腎臓は、背骨の両側にある2つの豆の形をした器官である。腎臓は血液からの老廃物を除去して尿を形成し、体内の水分と電解質の含有量を調節する。腎臓はタンパク質、グルコース、アミノ酸、重炭酸、及び無機リン酸を再吸収、保持する。そして血管内容積を維持するように作用するホルモンを放出する。腎臓は、体内の酸性/塩基のバランスの維持および血液pH電解質レベルと血圧の恒常性に寄与する。

【0004】

近位尿細管性上皮内に由来する腎細胞癌(RCC)は、成人の癌全体の約3%に相当する最も一般的なタイプの腎臓癌である。腎細胞癌には4つの主要なタイプがある。大部分は透明細胞タイプであり、顆粒細胞と透明細胞の混合タイプである。RCCは、非常に初期段階でのみ手術により治癒可能である。化学療法に対して耐性であり、また放射線治療に対してもかなりの耐性がある。RCCの発達に対する危険因子には喫煙と後天性の嚢胞腎疾患の発生がある。

10

【0005】

KSRCCをコードする哺乳動物のcDNAの発見は、腎臓の疾患、特に腎細胞癌の診断と治療に有用な組成物の提供により当分野の必要性を満たす。

【0006】

(発明の概要)

本発明は、後天性の嚢胞腎の疾患、特に腎細胞癌を含む腎臓の疾患の診断と治療に有用な、哺乳動物の腎臓特異タンパク質をコードする哺乳類cDNAの発見に基づく。

20

【0007】

本発明は、SEQ ID NO: 1のアミノ酸配列、SEQ ID NO: 1のアミノ酸配列に80%の同一性を有する変異体、SEQ ID NO: 1の抗原性エピトープ、SEQ ID NO: 1のオリゴペプチド、SEQ ID NO: 1の生物学的活性な部分を含む群から選択された哺乳動物タンパク質またはその部分をコードする、単離した哺乳動物のcDNAまたはその断片を提供する。さらに本発明は、SEQ ID NO: 2の核酸配列、SEQ ID NO: 2の核酸配列に83%の同一性を有する変異体、SEQ ID NO: 3-14の断片、SEQ ID NO: 2-14のオリゴヌクレオチドを含む群から選択された哺乳動物タンパク質またはその部分をコードする核酸配列を含む群から選択した単離した哺乳動物のcDNAまたはその相補体を提供する。更に本発明は、KSRCCをコードするcDNAまたはcDNAの相補体を含む組成物、基質、プローブを提供する。更に本発明は、そのようなcDNAを含むベクター、ベクターを含む宿主細胞、およびKSRCCを作製するためにcDNAを使用する方法を提供する。更に本発明は、KSRCCをコードするcDNAを含むベクターを有する遺伝形質転換細胞株または遺伝形質転換生物を提供する。本発明はまた、SEQ ID NO: 11-14から成る群から選択した哺乳動物cDNAの断片またはその相補配列を提供する。或る実施態様において、本発明はこれらの断片のうちの少なくとも1つを含む基体を提供する。2番目の実施態様において、本発明は検出、スクリーニング、精製方法で使用方法が可能で断片を含むプローブを提供する。更なる実施態様において、プローブは一本鎖相補的RNA及びDNA分子である。

30

40

【0008】

本発明は、核酸にハイブリダイズするプローブを含むサンプルの核酸の示差発現を検出するためにcDNAを使用する方法を提供する。それによりハイブリダイゼーション複合体を形成して、ハイブリダイゼーション複合体形成を標準と比較する。そして比較はサンプル中のcDNAの示差発現を示す。或る実施態様において検出方法は更にハイブリダイゼーション前にサンプルの核酸の増幅を含む。別の実施態様において、cDNAの示差発現を示す方法は腎細胞癌を診断するために使用される。別の実施態様において、cDNAまたは断片またはその相補配列はアレイの要素を含み得る。

【0009】

50

本発明はさらに、cDNAを特異結合する少なくとも1つのリガンドを同定するライブラリあるいは複数の分子または化合物をスクリーニングするためにcDNAまた断片またはその相補配列を使用する方法、そして特異結合を可能にする条件下でcDNAを分子または化合物に結合する、そしてcDNAへの特異結合を検出する、それによってcDNAを特異結合するリガンドを同定する方法を提供する。ある実施様態では、分子または化合物はアプタマー、DNA分子、RNA分子、ペプチド核酸、人工形成の染色体、ペプチド、転写因子、リプレッサー、調節分子から選択される。

【0010】

本発明は、SEQ ID NO: 1のアミノ酸配列、SEQ ID NO: 1のアミノ酸配列と少なくとも80%の相同性を有する変異体、SEQ ID NO: 1の抗原性エピトープ、SEQ ID NO: 1のオリゴペプチドまたSEQ ID NO: 1の生物学的活性な部分から成る群から選択した精製された哺乳動物タンパク質またはその部分を提供する。本発明は精製されたタンパク質または医薬用担体と同時にその部分から成る組成物も提供する。さらに本発明は腎細胞癌の対象を治療するためにKSRCCを使用する方法を提供する。それには治療の必要な患者に、精製されたタンパク質から成る組成物を投与することが含まれる。本発明はさらに、cDNAを特異結合する少なくとも1つのリガンドを同定するライブラリあるいは複数の分子または化合物をスクリーニングするためにタンパク質を使用する方法、そして特異結合を可能にする条件下でタンパク質を分子または化合物に結合する、そして特異結合を検出する、それによってタンパク質を特異結合するリガンドを同定する方法を提供する。ある実施様態においては、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、薬剤から選択される。別の実施様態では、腎細胞癌を患う患者を治療するためにリガンドを使用する。

【0011】

本発明は、被検サンプルの単離した抗体を含むタンパク質を特異結合する抗体用に被検サンプルをスクリーニングするために哺乳動物タンパク質を使用する、そして特異結合を可能にする条件下で単離した抗体をタンパク質に接触する、結合したタンパク質から抗体を解離する、そして抗体の存在または量が腎細胞癌の診断となる公知の標準と抗体の量を比較する方法を提供する。

【0012】

さらに本発明は、抗体を調製かつ精製するために、ある哺乳類タンパク質を使用する方法を提供し、その方法の内には抗体反応を誘発する条件下でそのタンパク質によって動物を免疫化すること、動物抗体を単離すること、ある基質にタンパク質を付着させること、そのタンパク質への特異結合を可能にする条件下でその基体を単離した抗体と接触させること、タンパク質から抗体を解離して精製した抗体を得ることが含まれる。

【0013】

本発明は、腎細胞癌で発現するタンパク質に特異的に結合する精製した抗体を提供する。本発明はまた、結合した抗体の量を公知の標準と比較して、抗体の結合を含む腎細胞癌を診断するために抗体を使用する方法を提供する。それによって腎細胞癌の存在を確定する。さらに本発明は腎細胞癌の患者を治療するために抗体を使用する方法を提供する。それには治療の必要な患者に、精製された抗体から成る医薬品成分を投与することが含まれる。

【0014】

本発明は、内在性ポリヌクレオチドの発現を防ぐためにマーカー遺伝子を哺乳動物のゲノムDNAに挿入する方法を提供する。さらに本発明は、哺乳動物モデル系を作製するためにcDNAを使用する方法を提供する。その方法の内には、SEQ ID NO: 2-14から選択されたcDNAを含むベクターを作製すること、そのベクターで胚性幹細胞を形質転換させること、形質転換した胚性幹細胞を選択すること、その形質転換した胚性幹細胞を哺乳動物胚盤胞に微量注入し、それによってキメラ胚盤胞を形成すること、キメラ胚盤胞を偽妊娠メスに導入させ、そのメスが生殖細胞にcDNAを含むキメラ子孫を産ませ

ること、そしてホモ接合体の哺乳動物モデル系を作製するためにキメラ哺乳動物を交配する方法が含まれる。

【0015】

(発明を実施するための形態)

本発明が、説明した特定の装置、材料及び方法に限定されるものではないことを理解されたい。また、ここで使用する専門用語は特定の実施例を説明する目的で用いたものであり、特許請求の範囲にのみ限定される本発明の範囲を限定することを意図したものではないことも併せて理解されたい。明細書中で用いている単数形の「或る」及び「その(この)」の表記は、文脈から明らかにそうでないとされる場合を除いて複数のものを指す場合もあることに注意しなければならない。例えば「或る宿主細胞」と記されている場合には当業者に公知であるがそのような宿主細胞が複数あることもある。

10

【0016】

本明細書中で用いる全ての専門用語及び科学用語は、特に定義されている場合を除き、当業者に一般に理解されている意味と同じ意味を有する。本発明で言及する全ての刊行物は、刊行物中で報告されていて且つ本発明に関係があるであろう細胞株、プロトコル、試薬及びベクターについて説明及び開示する目的で引用しているものである。本明細書のいかなる開示内容も、本発明が先行技術の効力によってこのような開示に対して先行する権利を与えられていないことを認めるものではない。

【0017】

定義

20

「KSRCC」は、天然、合成、半合成或いは組換え体など全ての哺乳動物種(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)から得られる実質的に精製されたタンパク質を指す。

【0018】

「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のcDNA体を表す。基質上の2から約40,000のcDNAの構成により、cDNAと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サイズとシグナル強度は確実に個別に区別できる。

【0019】

30

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

【0020】

「cDNA」とは、単離したポリヌクレオチド、核酸分子ある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に由来する、二本鎖または一本鎖である可能性があり、コードするまたはコードしない配列を表す、そしてゲノムDNA分子のイントロンを有するまたは有しないエキソンを表す。

【0021】

「タンパク質をコードするcDNA」という語は、当分野で公知である分析を使用することにより同定された保存された領域、モチーフある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された核酸配列を指す。これらの分析には保存された領域内で同一性を有する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)。

40

【0022】

「誘導体」とは化学修飾されたcDNAあるいはタンパク質を指す。cDNAの誘導体化にはクエオシン(queosine)ある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公知である。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

50

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、長い寿命あるいは強化された活性などの長所を付与する可能性がある。

【 0 0 2 3 】

「示差発現」とは、存在の有無、サンプル中の転写メッセンジャーRNAあるいは翻訳したタンパク質の量の少なくとも二倍の変更により検出される増加、または非調節、あるいは存在、あるいは減少、また下方調節または欠損、遺伝子発現を指す。

【 0 0 2 4 】

「疾患」とは、cDNAとKSRCCが他と異なって発現されている場合の症状、疾病、または症候群を指す。

【 0 0 2 5 】

「断片」とは長さが約200から約700塩基対の連続したヌクレオチドの鎖を指す。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で、またリガンドをスクリーニングするために結合アッセイで使用される可能性がある。核酸とそれらのリガンドはこのような方法で同定され、複製、転写または翻訳を調節する治療として有用である。

【 0 0 2 6 】

「Guilt by association (連帯責任により有罪)」(GBA)は、特定の疾患、調節経路、細胞内区分、細胞タイプ、組織の種類あるいは種と関連するcDNAまたはタンパク質を同定する方法である。特にこの方法は腎臓の疾患ための治療の診断、予後、治療、評価において有用なcDNAを同定する。

【 0 0 2 7 】

「ハイブリタイゼーション化合物」とは、1つの分子のプリンが3'-T-C-A-G-5'との5'-A-G-T-C-3'塩基対などの相補的な分子ピリミジンと水素結合する時、cDNAとサンプルの核酸の間で形成される相補性の度合およびヌクレオチド類似体の使用は、ハイブリダイゼーション反応の効率とストリンジェンシーに影響する。

【 0 0 2 8 】

「リガンド」とは、cDNA分子あるいはポリヌクレオチド上の相補部位あるいはタンパク質またはエピトープに特異結合する任意の物質、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あるいは調節し、核酸、タンパク質、糖質、脂肪、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。

【 0 0 2 9 】

「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または増幅技術、複製、転写または翻訳の調節で利用される可能性がある。同義語は、アンプライマー、プライマー、オリゴマーである。

【 0 0 3 0 】

「部分」とは、任意の目的に使用されるタンパク質の任意の部分に指す。しかし特にリガンドをスクリーニングするためまたは抗体を産生するためのエピトープを指す。

【 0 0 3 1 】

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、蛋白分解性切断及びその他が含まれ得る。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、酵素環境などによって異なる。

【 0 0 3 2 】

「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、in situ、ドットプロット法、アレイなどの技術またはスクリーニングアッセイを含む、ハイブリダイゼーション反応で使用するためにレポーター分子で標識化される。

【 0 0 3 3 】

「タンパク質」とは、ポリペプチドあるいはその任意の部分に指す。タンパク質の「部分

10

20

30

40

50

」とは、少なくとも1つの生物学的活性を保持するであろうアミノ酸配列の長さ、P F A MまたはP R I N T S分析によって同定されるドメインあるいはP R O T E A N プログラム(D N A S T A R , M a d i s s o n W I)のK y t e - D o o l i t t l e アルゴリズムを使用して同定したタンパク質の抗原性エピトープを指す。「オリゴペプチド」は、抗体を産生する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。

【0034】

「精製した」とは、自然環境から分離されたそして約60%から約90%まで自然に会合する他の化合物から遊離している任意の分子あるいは化合物を指す。

【0035】

「サンプル」は、核酸、タンパク質、抗体その他を含むような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、あるいは細胞から単離または抽出された膜、溶液中のまたは基板に固定されたゲノムDNA、RNA、またはcDNAと、細胞、組織プリント、フィンガープリント、口内細胞、皮膚あるいは髪などから成る可能性がある。

【0036】

「特異結合」とは、構造、特に分子側基に依存する、2つの分子間での特別で精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、2つの一本鎖核酸間のバックボーンに沿った水素結合あるい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。

【0037】

配列に適用される「類似性」とは、S m i t h - W a t e r m a n アルゴリズム(S m i t h 及びW a t e r m a n (1981) J M o l B i o l 147:195-197)あるいはB L A S T 2 (A l t s c h u l r a (1997) N u c l e i c A c i d s R e s 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あるいは残基の定量化(通常は%)を指す。B L A S T 2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

【0038】

「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、またはその他の管、プレート、ポリマー、微細粒子が含まれ、穴、溝、ピン、チャンネル、孔等、様々な表面形態を有する。

【0039】

「変異体」とは、cDNAまたはcDNAがコードするタンパク質の認識される変異である分子を指す。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、B L A S Tスコアにより決定する可能性がある。対立遺伝子変異体はcDNAに対して高一致率を有し、100塩基に付き約3塩基が異なる。「1塩基多型性」(S N P)とは、欠失、挿入または置換による単一の塩基での変異を指す。変異は保存(プリンからプリン)あるいは非保存(プリンからピリミジン)される可能性があり、コード化されたアミノ酸または2次、3次、4次構造の変異となる可能性もならない可能性もある。

【0040】

(発明)

本発明は、K S R C CをコードするcDNAの発見と、腎臓の疾患の性質決定、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。

【0041】

腎臓経路で同時に調節される配列のG B A分析によって、本発明のK S R C Cをコ - ド化

10

20

30

40

50

する核酸が腎臓特異的として最初に同定化された（腎臓組織 cDNA ライブラリ、K I D N N O T 3 1 からの I n c y t e クローン 3 4 8 1 9 4 2 C B 1）。腎臓病と関連すると知られている 11 の遺伝子を選択して、腎臓の疾患と密接に関連する新規な遺伝子を同定した。これらの既知の遺伝子は、u r o m o d u l i n、N K C C 2、N C C T、アルドラーゼ B、R O M K 1、A T P 1 G 1、P D Z K 1、N P T 1、カルピンジン、キニノゲン、C I C K b である。0 . 0 0 0 0 1 未満のカットオフ p 値を使用して、確率値によって関連の度合いを測定した。配列をさらに検証して、確率テストを通過した遺伝子が既知の腎臓病と関連する遺伝子と強い関連性をもつことを確実にした。新規な 9 つの遺伝子が、総計 4 1 , 4 1 9 から成る既知の腎臓病と関連する遺伝子と強い結合を示した。9 つの遺伝子の各々は p 値 $10e^{-5}$ 未満で 11 の既知の遺伝子のうちの少なくとも 1 つと同時に発現した。K S R C C cDNA は長さが 2 0 5 4 の核酸である。S E Q I D N O : 3 (I n c y t e クローン 3 4 8 1 9 4 2 C B 1) は長さが 5 2 1 ヌクレオチドであり、B L A S T 2 を使用して K S の核酸配列、ラット腎臓特異的タンパク質 (g 3 1 2 7 1 9 2) と 8 2 % の同一性を示す。

【0042】

S E Q I D N O : 2 の cDNA は下記の重複した核酸配列および/または伸長した核酸配列に由来する。(S E Q I D N O : 3 - 1 0) : I n c y t e クローン 3 4 8 1 9 4 2 C B 1、5 5 1 9 4 2 7 H 1 (L I V R D I R 0 1)、5 3 9 0 9 8 4 F 8 (K I D N N O T 3 2)、4 3 3 2 2 3 0 T 9 (K I D N N O T 3 2)、7 6 0 8 3 5 R 1 (B R A I T U T 0 2)、7 6 2 5 4 0 R 1 (B R A I T U T 0 2)、2 0 9 9 4 2 0 R 6 (B R A I T U T 0 2) 及び 7 6 9 5 9 0 5 J 1 (L G c o m p s e q s J U N 2 0 0 0 n e w)。

【0043】

図 3 は腎臓組織、特に腎細胞癌を患う患者の組織における転写物の発現を示す。よって、cDNA は腎細胞癌を診断するアッセイに有用である。ヌクレオチド 1 からヌクレオチド 106 までの S E Q I D N O : 2 の断片は、ハイブリダイゼーションプローブとして有用である。54 から 69 あるいは 80 から 95 のオリゴヌクレオチドは、他の腎臓特異タンパク質から K S R C C をコードする転写物を区別する診断として有用である。

【0044】

一実施態様では、本発明は図 1 A から 1 F で表示されるようにアミノ酸配列 S E Q I D N O : 1 を含むポリペプチドを提供する。K S R C C は長さが 577 アミノ酸であり、B L A S T 2 分析を使用してラット腎臓特異的タンパク質 (g 3 1 2 7 1 9 3) のアミノ酸配列と 77 % の同一性を示す。B L I M P S 分析は、S E Q I D N O : 1 には 1 つの 16 アミノ酸セグメントがあり、それが A M P 結合ドメイン g 3 1 2 7 1 9 3 と一致することを示す。B L O C K S 分析は、Q 2 1 3 から T 2 2 4 と S 2 2 5 から H 2 3 3 の K S R C C の領域が A M P 結合部位と類似であることを示す。P f a m 分析は、N 8 2 から V 4 9 3 の K S R C C の領域が A M P 結合部位と類似であることを示す。

【0045】

図 2 A と 2 B が示すように、K S R C C はラットの腎臓特異タンパク質 (g 3 1 2 7 1 9 3 : S E Q I D N O : 1 5)、ラットのタンパク質のヒト相同体 ((g 3 2 1 9 3 3 9 : S E Q I D N O : 1 6) と化学的及び構造的類似性を示す。特に K S R C C とラットの腎臓特異タンパク質は、およそ同じ長さのタンパク質に対して 77 % の配列同一性を共有する (それぞれ 577 アミノ酸残基と 572 アミノ酸残基)。ラットのタンパク質のヒト相同体 (g 3 2 1 9 3 3 9) は、より短いタンパク質 (207 アミノ酸残基) であるが、この全長に対して K S R C C に 97 % 同一である。

【0046】

図 4 A と 4 B が示すように、これらの 3 つのタンパク質は類似の親水性プロットと抗原指数を示す。

【0047】

K S R C C をコードする哺乳動物の cDNA の変異体は、デフォルトパラメータで B L A

ST2とZOOSEQデータベース(Incyte Genomics)を使用して同定された。KSRCをコードする哺乳動物のcDNAの変異体には、それぞれ配列表の701648693H1(RALITXT40)、212451_Rn.1(鋳型)、202264_Rn.1(鋳型)、701836346T1(RAKITXT11)、SEQ ID NO: 11-14がある。これらのcDNAは、ヒト腎臓疾患をモデルとし、また腎臓疾患のための潜在的な治療上の処置を検査することが可能である遺伝形質転換細胞株または遺伝形質転換生物を作製するために特に有用である。下記の表1はKSRCをコードするcDNAとその哺乳動物変異体、SEQ ID NO: 11-14の間の一致率を示す。列1はヒトcDNAのSEQ ID、列2は変異体cDNAのSEQ ID var、列3は変異体cDNAのクローン番号、列4はヒトcDNAへの一致率、列5はヒトcDNAへの変異体cDNAのアラインメントを示す。

SEQ ID var	SEQ ID var	クローン var	一致	NtH アラインメント
2	11	701648693H1	83%	407-513
2	12	212451_Rn.1	84%	709-986
2	13	202264_Rn.1	83%	1067-1764
2	14	701836346T1	84%	1655-1812

10

【0048】

これらのcDNAは、ヒト腎臓疾患をモデルとし、また腎臓疾患のための潜在的な治療上の処置を検査することが可能である遺伝形質転換細胞株または遺伝形質転換生物を作製するために特に有用である。

20

【0049】

当業者であれば、遺伝暗号の縮重の結果、KSRCをコードする多数のcDNA(既知の遺伝子または天然の遺伝子のcDNAと最低限の類似性しか有しないcDNAもある)を産出し得ることは理解できよう。したがって本発明には、可能コドン選択に基づく組合せの選択によって産出し得るようなありとあらゆる可能性のあるcDNAの変異体を網羅し得る。これらの組み合わせは、天然のKSRCをコード化するポリヌクレオチドに適用される標準的なトリプレット遺伝暗号を基に作られ、全ての変異が明確に開示されていると考慮する。

30

【0050】

SEQ ID NO: 2-14のcDNAとその断片は、SEQ ID NO: 2とサンプル中の関連する分子の中で同定、区別するためのハイブリダイゼーション、増幅、スクリーニング技術で使用可能である。哺乳動物cDNAは、ヒトの腎細胞癌のモデル系である遺伝形質転換細胞株または遺伝形質転換生物を作製するために使用することが可能であり、潜在的な治療上の処置の薬剤効果および毒性を検査することが可能である。cDNA、タンパク質、抗体、分子、本発明のcDNA sとタンパク質を使用して同定された化合物を使用して、毒性研究、臨床試験、被検者/患者治療プロフィールを実施すること、モニターすることが可能である。

40

【0051】

本発明の性質決定と使用

G B A 分析

G B Aは、複数のcDNAライブラリで発現するcDNAを同定する。cDNAは、特定の疾患の過程、細胞内区分、細胞タイプ、組織の種類あるいは種で発現する既知または未知の機能の遺伝子を含む。既知の機能の遺伝子の発現パターンを未知の機能のcDNAのそれと比較することにより、特異な同時発現確率閾値を満たすかどうかを決定する。この比較により、既知の遺伝子での同時発現確率の高いcDNAのサブセットを同定することが可能となる。高い同時発現確率は、特定の同時発現確率閾値と相関し、その閾値は好ましくは0.001未満、更に好ましくは0.00001未満である。

50

【0052】

cDNAは、多様な源泉に由来するcDNAライブラリから始まる。限定されるものでは

ないが多様な源泉には、ヒト、マウス、ラット、イヌ、サル、植物、酵母等の真核生物、バクテリア、ウイルス等の原核生物を含む。これらのcDNAは多様な配列の種類から選択することができる。限定するものではないが、発現配列タグ(EST)、構築したポリヌクレオチド、完全長遺伝子コード領域、プロモーター、イントロン、エンハンサー及び5'及び3'の非翻訳領域を含む。統計的に有意義な分析結果を得るために、少なくとも5つのcDNAライブラリでcDNAを発現する必要がある。

【0053】

本発明の同時発現分析で使用されるcDNAライブラリは、任意の細胞または細胞株、組織または器官から得ることができる。また副腎、胆道、膀胱、血球、血管、骨髄、脳、気管、軟骨、クロム親和系、結腸、結合組織、培養細胞、胚幹細胞、内分泌腺、上皮組織、食道、胎児、神経節、心臓、視床下部、免疫機構、腸、ランゲルハンス島、腎臓、咽頭、肝臓、肺、リンパ、筋肉、神経、卵巣、膵臓、陰茎、末梢神経系、食細胞、下垂体、胎盤、胸膜、前立腺、唾液腺、精嚢、骨格、脾臓、胃、精巣、胸腺、舌、尿管、子宮等からも可能である。選択されたcDNAライブラリの数の範囲は、わずか500から10,000以上までの可能性がある。

10

【0054】

腎臓の疾患と関連のある病気の診断あるいは予後マーカーとしてあるいは治療標的としての遺伝子の使用に基づいて、既知の腎臓特異遺伝子を選択することができる。

【0055】

既知の腎臓特異遺伝子との統計的に有意義な同時発現パターンを示すcDNAを同定する方法は下記の通りである。最初に、cDNAライブラリの遺伝子配列の存在あるいは不存在を定義する。ある遺伝子に対応する少なくとも1つのcDNA断片がライブラリから採取したcDNAサンプルで検出される時、その遺伝子はcDNAライブラリに存在する。そして対応するcDNA断片がサンプル中で検出されない時、その遺伝子はライブラリに存在しない。

20

2番目に、確率方法を用いて遺伝子同時発現の有意性を評価して、同時発現の偶然による確率を測定する。確率方法はフィッシャーの正確確率検定、カイ2乗検定またはカップパ検定である可能性がある。これらの試験と適用の例は公知であり、標準的な統計学のテキストで見出すことができる(Agresti (1990) Categorical Data Analysis, John Wiley & Sons, New York NY; Rice (1988) Mathematical Statistics and Data Analysis, Duxbury Press, Pacific Grove CA)。ボンフェローニ検定(前出のRice、384ページ)も、1つの遺伝子対複数の他の遺伝子の統計結果を修正するために確率方法の1つと併用して適用することができる。好適な実施例では、偶然による確率をフィッシャーの正確確率検定により測定する。また偶然による確率の閾値を好ましくは0.001未満、更に好ましくは0.00001未満に設定する。

30

【0056】

2つの遺伝子AとBが類似の同時発現のパターンを有するかどうかを決定するために、出現性データベクターを下記の表で説明されているように產生する。ライブラリで少なくとも1回出現する遺伝子の存在を1と示す。またライブラリでの不存在を0と示す。

40

	ライブラリ 1	ライブラリ 2	ライブラリ 3	...	ライブラリ N
遺伝子 A	1	1	0	...	0
遺伝子 B	1	0	1	...	0

10

【 0 0 5 7 】

所定の対の遺伝子には、出現性データを 2 × 2 分割表で概略することができる。

	遺伝子 A 存在	遺伝子 A 不在	合計
遺伝子 B 存在	8	2	10
遺伝子 B 不在	2	18	20
合計	10	20	30

20

30

【 0 0 5 8 】

分割表は、総計 30 ライブラリに及ぶ遺伝子 A と遺伝子 B の同時出現データを示す。遺伝子 A と遺伝子 B の両方はライブラリで 10 回出現しており、この表は下記を概略して示している。(1) 遺伝子 A と B 両者がライブラリで存在する回数、(2) 遺伝子 A と B 両者がライブラリで不在の回数、(3) 遺伝子 A が存在して、遺伝子 B が不在の回数、(4) 遺伝子 B が存在して、遺伝子 A が不在の回数上左エントリはライブラリで 2 つの遺伝子が同時出現する回数であり、中右エントリはライブラリで両者の遺伝子が出現しない回数である。非対角エントリは 1 つの遺伝子が出現するが別の遺伝子が出現しない回数である。A と B の両者が存在するのは 8 回であり、不在するのは 18 回である。遺伝子 A が存在して B が不在の場合が 2 回、遺伝子 B が存在して遺伝子 A が不在の場合は 2 回であるフィッシャーの正確確率検定を利用した計算によると、上記の関連性が偶然によって起こった確率(「p 値」)は、0.0003 である。p 値が 0.01 未満である場合、関連性は通常有意と考慮される(前出の Agrest i と Rice)。

40

【 0 0 5 9 】

2 つの遺伝子の同時出現の確率を評価する方法には幾つかの前提がある。方法はライブラリが独立しており、同様にサンプル化していると想定する。しかし実際の状況では、選択

50

した cDNA ライブラリは完全には独立していない。1 つ以上のライブラリを単一の対象または組織から得ることが可能だからである。また完全には同様にサンプル化されてもいない。異なる数の cDNA を各ライブラリから配列化する可能性があるからである。通常シーケンスした cDNA の数の範囲は、ライブラリ毎に 5,000 から 10,000 の cDNA である。更に、フィッシャーの正確同時発現確率を各遺伝子対少なくとも 5 つのライブラリで出現する他のすべての構築した遺伝子に対して計算する。また複数の統計試験のためにボンフェローニ検定を用いた。

【0060】

cDNA ライブラリ

本明細書で開示する特定の実施例において、mRNA を当業者に周知の方法を使用して哺乳動物細胞と組織から単離して、cDNA ライブラリを準備するために使用した。上記に示された Incyte クローンを哺乳動物 cDNA ライブラリから単離した。本発明の代表的な 3 つのライブラリ調製が下記の実施例に記載されている。コンセンサス配列は、化学的そして / あるいは電子的に Incyte cDNA を含む断片そして PHRAP (P Green, University of Washington, Seattle WA) と AUTOASSEMBLER application (Applied Biosystems, Foster City CA) などのコンピュータプログラムを使用して伸長そして / あるいはショットガン配列から構築された。クローン、伸長そして / あるいはショットガン配列をクラスターそして / あるいはマスタークラスターに電子的に構築した。

10

20

【0061】

シーケンシング

核酸をシーケンシングする方法は当分野でよく知られており、本発明の何れの実施例も核酸をシーケンシングする方法を用いて実施可能である。核酸をシーケンシングする方法には酵素を用いることができる。例えば DNA ポリメラーゼ I のクレノウ断片、SEQUENASE、Taq DNA ポリメラーゼ、熱安定性 T7 DNA ポリメラーゼ (Amersham, Pharmacia Biotech (APB), Piscataway NJ) を用いることができる。或いは、例えば ELONGASE 増幅システム (Life Technologies, Gaithersburg MD) において見られるように、ポリメラーゼと校正エキソヌクレアーゼを併用することができる。好適には、MICROLAB 2200 液体転移システム (Hamilton, Reno, NV)、DNA ENGINE サーマルサイクラー (MJ Research, Watertown MA) 等の装置を用いて配列の調製を自動化する。シーケンシングをするために一般的には使用される装置は、ABI PRISM 3700、377 或いは 373 DNA シーケンシングシステム (Applied Biosystems)、MEGABACE 1000 DNA シーケンシングシステム (APB) などである。配列は、当分野で公知であり、Ausubel, ら (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7、Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, 856 - 853 ページに記載されている、多種のアルゴリズムを使用して分析され得る。

30

40

【0062】

ショットガン・シーケンシングは、目的の特定クローン化インサートの配列を完成するために使用することもできる。ショットガン・ストラテジーには、無作為に元来のインサートを多様なサイズのセグメントに切断すること、これらの断片をベクターにクローニングすることが含まれる。これらの断片は、元来のインサートの配列全体が知られるまで重畳する末端を使用して、配列化され、再構築される。ショットガン・シーケンシング方法は当分野で公知であり、目的の cDNA s に隣接する代表的領域から選択される熱安定性 DNA ポリメラーゼ、易熱性 DNA ポリメラーゼ、プライマーを使用する。不完全に構

50

築された配列は、当分野で公知である多様なアルゴリズムまたはCONSED (Gordon (1998) Genome Res 8:195-202)などのプログラムを使用して同一性を調べられる。ベクターまたはキメラ配列を含む汚染された配列あるいは除去された配列は、それぞれ除去、回復が可能であり、不完全に構築された配列を完全な配列に組織化できる。

【0063】

核酸の伸長

本発明の配列は、当分野で周知の多様なPCR法をベースにした種々の方法で伸長することが可能である。例えば、XL-PCRキット(Applied Biosystems)、ネステッドプライマー、市販のcDNAあるいはゲノムDNAライブラリは核酸を伸長するために使用され得る。全てのPCRベースの方法に対して、市販されているソフトウェア、例えばOLIGOプライマー分析ソフトウェア(Molecular Biology Insights, Cascade CO)などを用いて、長さが約22~30ヌクレオチド、GC含有率が約50%以上、温度約55~68で標的分子に対してアニーリングするようにプライマーを設計し得る。調節エレメントを回収するために配列を伸長する時、cDNAライブラリよりもゲノムを使用するほうが好適である。

10

【0064】

ハイブリダイゼーション

cDNAとその断片は、様々な目的でハイブリダイゼーション技術において使用することが可能である。プローブを5'調節領域などの固有の領域または非保存された領域(即ちタンパク質の保存された触媒ドメインをコードするヌクレオチドの5'または3')から由来する、また設計することは可能である。そしてKSRCC、対立遺伝子変異配列あるいは関連分子をコードする天然の分子を同定するために、プロトコルで使用する。プローブはDNAまたはRNAそして一本鎖である可能性があり、任意の核酸配列SEQ ID NO: 2-14に対して少なくとも50%の配列同一性を有するはずである。ハイブリダイゼーションプローブはオリゴ標識化、ニックトランスレーション、末端標識化、またはレポーター分子の存在下でPCR増幅を使用して産生され得る。cDNAまたはその断片を含むベクターは、RNAポリメラーゼと標識されたヌクレオチドを加えてin vitroでmRNAプローブを産生するために使用することが可能である。このような方法は、例えばAPBから市販されている種々のキットを用いて実行することができる。

20

30

【0065】

ハイブリダイゼーションのストリンジェンシーは、プローブのG+C含有、塩分濃度、温度により決定する。特に、ストリンジェンシーは、塩の濃度の減少あるいはハイブリダイゼーション温度の上昇により増強することができる。膜系ハイブリダイゼーションで使用する溶液では、ホルムアミドなどの有機溶剤の添加により低温で反応が起きることが可能となる。ハイブリダイゼーションは、1%のドデシル硫酸ナトリウム(SDS)を60で5xSSCなどのバッファーで低ストリンジェンシーで実施することができる。そして不一致を含む核酸配列の間でハイブリダイゼーション化合物の形成を可能にする。続いて0.1% SDSを45(中間ストリンジェンシー)または68(高ストリンジェンシー)で、0.2xSSCなどのバッファーで高いストリンジェンシーで洗浄を実施した。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核酸が完全に相補的な場合のみ安定を保持するであろう。或る膜系ハイブリダイゼーションでは、好ましくは35%または最も好ましくは50%、ハイブリダイゼーションが実施される温度を低下させるためにホルムアミドはハイブリダイゼーション溶液に加えることができる。そしてバックグラウンドシグナルはSarkosylまたはTRITON X-100(Sigma-Aldrich, St. Louis MO)などの他の界面活性剤、変性したサケ精子DNAなどのブロッカーの使用により減少させることができる。化合物の選択とハイブリダイゼーションの条件は当業者には公知である。またAusubel(前出)およびSambrookら(1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, P

40

50

l a i n v i e w N Y に概説されている。

【0066】

アレイを本技術分野でよく知られている方法を用いて調製し、そして分析することができる。オリゴヌクレオチドはアレイでプローブあるいは標的として使用され得る。アレイは、同時に多数の遺伝子の発現レベルをモニターするためにまたは遺伝変異体、突然変異及び単一ヌクレオチド多型性を同定するために使用され得る。そのような情報を用いることで、遺伝子機能を決定し、症状、疾患または障害の遺伝的根拠を理解し、症状、疾患または障害を診断し、薬剤の活性を開発及びモニターすることができる。(Brennan, ら (1995) の米国特許第5,474,796号、Schena, ら (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619、Baldeschweiler らの (1995) PCT出願第WO95/251116号、Shallon, らの (1995) PCT出願第WO95/35505号、Heller, ら (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155、Heller, M. J. らの (1997) 米国特許第5,605,662号等を参照)。

10

【0067】

ハイブリダイゼーションプローブは、天然のゲノム配列をマッピングするのにも有用である。プローブを(1)特定の染色体、(2)染色体の特定領域または、(3)人工形成の染色体、例えば、ヒト人工染色体(HAC)、酵母人工染色体(YAC)、細菌人工染色体(BAC)、細菌P1産物、或いは単一染色体cDNAライブラリに対してハイブリダイズすることが可能である。

20

【0068】

発現

KSRCCをコードする多数のcDNAのいずれかは、ベクターにクローニングされ得る。また宿主細胞でタンパク質或いはその部分を発現するために使用される。核酸配列は、新規な制限酵素部位の作製、グリコシル化パターンの変更、コドン優先の変更のためにDNAシャッフリング(USPN 5,830,721)、特定部位の突然変異誘発などの方法により操作することが可能であり、特定の宿主における発現の増大、スプライス変異体の生成、半減期の延長などがなされる。発現ベクターは、特定の宿主での効率のために選択された多様な源泉からの転写調節エレメント及び翻訳調節エレメント(プロモーター、エンハンサー、特定の開始シグナル、ポリアデニル化された3'配列)を含むことが可能である。ベクター、cDNA、調節エレメントは、当分野で公知であり、Sambrook (前出4、8、16、17章)に記載されている、in vitro組換えDNA技術、合成技術、及びin vivo遺伝子組換え技術を利用して結合する。

30

【0069】

種々の宿主系が、発現ベクターで形質転換が可能である。限定するものではないがこのような宿主系には、組換えバクテリオファージ、プラスミドまたはコスミドDNA発現ベクターで形質転換させた細菌や、酵母菌発現ベクターで形質転換させた酵母菌や、バキュロウイルス発現ベクターで形質転換させた昆虫細胞系や、ウイルスそして/あるいは細菌性エレメントを含む発現ベクターで形質転換させた植物細胞系、あるいは動物細胞系がある(前出のAusubel, unit 16)。例えば、アデノウイルス転写物/翻訳複合体は哺乳動物細胞で利用することが可能である。配列をウイルスのゲノムのE1あるいはE3領域にライゲーションさせた後に、感染ウイルスは宿主細胞のタンパク質を形質転換、発現するために用いられる。ラウス肉腫ウイルスエンハンサーあるいはSV40またはEBVをベースにしたベクターを用いて、タンパク質を高レベルで発現させることもできる。

40

【0070】

核酸配列の日常的なクローニング、サブクローニング、増殖には、多機能のPBLUESCRIPT(Stratagene, La Jolla CA)またはpSPORT1プラスミド(Life Technologies)を用いることができる。これらのベク

50

ターの多数のクローニング部位に核酸配列を導入すると *lacZ* 遺伝子が破壊され、形質転換された細菌用の比色スクリーニング法が可能となる。更にこれらのベクターは、クローニングされた配列における *in vitro* 転写、ジデオキシのシーケンシング、ヘルパーファージによる一本鎖の救出、入れ子状態の欠失の生成にも有用であろう

【0071】

組換えタンパク質の長期にわたる産生のためには、同じ或いは別のベクターの上の選択可能マーカーまたは可視マーカー遺伝子と共にベクターは安定して細胞株に形質転換されることが望ましい。ベクターの形質転換後、細胞は選択培地に移す前に強化培地で約1~2日間細胞を増殖させることができる。選択可能マーカー、代謝拮抗物質、抗生物質或いは除草剤遺伝子は、関連する選択培地へ耐性を与え、導入された配列をうまく発現するような細胞の成長及び回収が可能となる。アントシアニン、緑色蛍光タンパク質(GFP)、βグルクロニダーゼ、ルシフェラーゼ等の選択的培地上の生存あるいは可視マーカーの発現により同定された耐性クローンは、培養技術を用いて増殖可能である。可視マーカーは、導入された遺伝子により発現したタンパク質の量を定量するためにも用いられる。宿主細胞が望ましい哺乳動物cDNAを含んでいるかの検証は、DNA-DNAまたはDNA-RNAハイブリダイゼーションあるいはPCR増幅技術に基づく。

10

【0072】

宿主細胞株は、所望の形に組換えタンパク質を調節する能力によって選択がなされ得る。そのような修飾には、アセチル化、カルボキシル化、グリコシル化、リン酸化、脂質化及びアシル化及びその他がある。「プレプロ」形を切断するような翻訳後処理を利用して、タンパク質のターゲティング、折りたたみ及び/または活性を特定することも可能である。翻訳後の活性のための固有の細胞装置及び特徴のある機構を有する種々の宿主細胞は、American Type Culture Collection(ATCC, Manassas VA)から入手可能であり、外来タンパク質の正しい修飾及び処理を確実にするように選択し得る。

20

【0073】

細胞培地でのこのタンパク質の回収

精製を容易にするためのベクターに組み換えられる異種部分には、グルタチオンSトランスフェラーゼ(GST)、6xHis、フラグ、MYC及びその他がある。GSTは固定化グルタチオン上で、そして6-Hisは金属キレート樹脂上で、市販されている親和性基質を用いて精製する。FLAGとMYCは、市販されているモノクローナル及びポリクローナル抗体を用いて精製する。精製後の分離を容易にするために、タンパク質と異種部分の間に位置するベクターにタンパク質分解切断部位をコードする配列を含めることもできる。組換えタンパク質の発現及び精製の方法は、前出のAusubel(1995)16章に記載されており、市販されている。

30

【0074】

ペプチドの化学合成

タンパク質あるいはその部分は、組換え方法だけでなく、当分野で公知の化学的方法によっても産生され得る。固相ペプチド合成は、バッチごとの処理または連続的な流れ処理ででき、α-アミノと側鎖によって保護されるアミノ酸残基をリンカー群を通して不溶性重合体支持に順次加えて実施され得る。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群は、支持樹脂を形成するためにスチレン-ジビニルベンゼン共重合体に着する。アミノ酸残基は、酸に不安定なBoc(t-ブチルオキシカルボニル)または塩基に不安定なFmoc(9-フルオレニルメトキシカルボニル)によりN-α保護(N-α-protected)される。保護アミノ酸のカルボキシル基は、残基を固相支持樹脂にアンカーするリンカー群のアミンに結合する。トリフルオロ酢酸またはピペリジンは、それぞれBocまたはFmocについては保護基を除去するために用いられる。各追加のアミノ酸を、カップリング剤が前もって活性化されたアミノ酸誘導体を用いてアンカーした残基に加えて、次いで樹脂を洗浄する。全長ペプチドを連続脱保護、アミノ酸誘導体の結合、ジクロロメタンまたは/あるいはN,N-ジメチルホルムアミドでの洗浄に

40

50

より合成する。前記ペプチドを、ペプチド酸またはアミドを産生するためにペプチドカルボキシル末端とリンカ - 群の間で切断する。(Novabiochem 1997/98 Catalog and Peptide Synthesis Handbook, San Diego CA S1-S20ページ) 自動合成はABI 431Aペプチドシンセサイザ(Applied Biosystems)を用いて実施し得る。タンパク質或いはその部分は、分離用高速液体クロマトグラフィーを用いて精製可能である。またその組成物はアミノ酸分析かシーケンシングにより確認することができる(Creighton (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY)。

【0075】

10

抗体の調製とスクリーニング

ヤギ、ウサギ、ラット、マウス、ヒト、その他を含む種々の宿主は、KSRCC、またはKSRCCの任意の断片を注入することによって免疫化することができる。フロイントアジュバント、ミネラルゲルアジュバント等のアジュバント、またリゾレシチン、ブルロニックポリオール、ポリアニオン、ペプチド、油性乳剤、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等の界面活性剤とがあり、免疫反応を増大するために用いることも可能である。抗体を誘発するために用いられるオリゴペプチド、ペプチドまたはタンパク質の部分は、少なくとも約5のアミノ酸からなり、天然のタンパク質の部分に同一の約10のアミノ酸からなるものが好ましい。オリゴペプチドは、KLHなどの別のタンパク質と融合し、キメラ分子に対する抗体が産生され得る。

20

【0076】

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

【0077】

別法では、当分野で周知の方法を用いて、一本鎖抗体の産生のための記載された技術を適用して、エピトープ特異性一本鎖抗体を生成する。タンパク質のエピトープに対する特異的な結合部位を含む抗体断片も得ることができる。例えば、限定するものではないが、このような断片には、抗体分子のペプシン消化によって作製されるF(ab')₂断片と、F(ab')₂断片のジスルフィド架橋を還元することによって作製されるFab断片とがある。或いは、Fab発現ライブラリを作製することによって、モノクローナルFab断片を所望の特異性と迅速且つ容易に同定することが可能となる(Huseら(1989) *Science* 246:1275-1281等を参照)。

30

【0078】

KSRCCまたはその任意の部分は、所望の特異性を有する抗体を同定するphagemidまたはB-リンパ球免疫グロブリンライブラリのスクリーニングアッセイで用いることができる。確立された特異性を有するポリクローナル抗体またはモノクローナル抗体の何れかを用いる競合的な結合、または免疫学的検定のための数々のプロトコルが、当分野では周知である。通常このような免疫学的検定には、タンパク質とその特異性抗体との間の複合体形成の計測が含まれる。2つの非干渉エピトープに反応するモノクローナル抗体を用いた、2部位のモノクローナルベース免疫アッセイ(two-site, monoclonal-based immunoassay)が好ましいが、競合の結合アッセイも用いることもできる(Pound(1998) Immunochemical Protocols, Humana Press, Totowa NJ)。

40

【0079】

アッセイのための分子の標識化

50

多岐にわたるレポーター分子及び結合方法が、当業者に知られており、様々な核酸アッセイ、アミノ酸アッセイおよび抗体アッセイにこれらの方法を用い得る。標識した分子の合成は、 $^3\text{ }^2\text{P}$ -dCTP (APB)、Cy3-dCTPまたはCy5-dCTP (Operon Technologies, Alameda CA)などの標識されたヌクレオチドあるいは $^3\text{ }^5\text{S}$ -メチオニン (APB)などのアミノ酸を取り込むための市販のキット (Promega, Madison WI)を用いて達成することが可能である。ヌクレオチドとアミノ酸は、BIODIPYまたはFITC (Molecular Probes, Eugene OR)等の試薬を用いて分子内に存在するアミン、チオールその他の基への化学結合により蛍光剤、化学発光剤、発色剤及びその他を含む多種の物質を用いて直接標識化することが可能である。

10

【0080】

(診断)

cDNA、断片、オリゴヌクレオチド、相補的RNA及びDNA分子そしてPNAは、遺伝子の示差発現、mRNAの不在、存在及び過剰発現を検出し定量するために、そして治療インターベンション中にmRNAレベルをモニターするために用いることができる。同様に、KSRCCを特異結合する抗体は、タンパク質を定量するために用いられる。示差発現と関連する腎臓の疾患には、後天性の嚢胞腎の疾患と、特に腎細胞癌を含む。診断アッセイは、示差遺伝子発現を検出するために患者の生物学的サンプル中の遺伝子発現を基準サンプルと比較する、ハイブリダイゼーションまたは増幅技術を用い得る。この比較のための定性方法または定量方法は、当分野で公知である。

20

【0081】

例えば、cDNAあるいはプローブは標準的な方法で標識化され、ハイブリダイゼーション複合体の形成に好適な条件下で患者の生物学的サンプルに加える。インキュベーション期間後、サンプルを洗浄し、ハイブリダイゼーション複合体と関連する標識 (またはシグナル) の量を定量して標準値と比較する。正常あるいは病変値と比較して患者サンプルの複合体の形成が著しく変化している場合 (高いまたは低い) は、示差発現は疾患の存在を示している。

【0082】

示差発現を確立する基準となるものを提供するために、正常と疾患の発現プロファイルを確立する。これは、発生するハイブリダイゼーションに好適な条件下で、cDNAと動物

30

【0083】

このようなアッセイは、動物実験、臨床試験における特定の治療効果を推定するため、或いは個々の患者の治療をモニターするために用いることもできる。症状の存在を一旦確定して治療プロトコルを開始すると、診断アッセイを通常ベースで繰り返して、患者の発現

40

【0084】

免疫学的方法

特異的なポリクローナル抗体またはモノクローナル抗体のどちらかを用いるタンパク質の発現の検出及び計測のための免疫学的方法は、当分野で周知である。このような技法には、酵素に結合した免疫吸着剤検定法 (ELISA)、ラジオイムノアッセイ (RIA)、フローサイトメーター (FACS) などがある。2つの非干渉エピトープに反応するモノクローナル抗体を用いた、2部位のモノクローナルベースイムノアッセイ (two-site, monoclonal-based immunoassay) が好ましいが

50

、競合の結合アッセイも用いることもできる。(Coligan ら (1997) Current Protocols in Immunology, Wiley-Interscience, New York NY、前出 Pound を参照)。

【0085】

(治療)

図2Aと2Bが示す通り、AMP結合ドメインに関して、KS RCC (SEQ ID NO: 1) とラットとヒトの腎臓特異タンパク質(それぞれg312719、g321933)の間で化学的及び構造的類似性が存在する。さらに図3で示されているように示差発現が組織そして腎細胞癌に強く関連している。KS RCCは明らかに後天性の囊胞腎の疾患、特に腎細胞癌に関与している。

10

【0086】

腎細胞癌などのタンパク質の発現の増大に関連する症状の治療においては、発現またはタンパク質活性を低下させることが望ましい。一実施例では、タンパク質のインヒビター、アンタゴニストまたは抗体を患者に投与して、発現または活性の増大に関連した症状を治療することも可能である。別の実施例では、内在性タンパク質の発現または活性の増大に関連した症状の治療のために、インヒビター、アンタゴニストまたは抗体を含む医薬品組成物を医薬用担体と共に患者に投与することも可能である。別の実施例では、症状の治療のために、cDNAまたはその断片の相補配列を発現するベクターを患者に投与することも可能である。

【0087】

20

cDNA、相補的分子またはその断片、タンパク質またはその部分、これらの核酸分子を送達するまたはタンパク質を発現するベクターそしてそれらのリガンドの何れかを他の薬剤と共に投与することは可能である。併用療法で用いる治療薬は、当業者が従来の医薬原理に従ってを選択し得る。治療薬と組み合わせることにより、各薬剤をより低い投与で特定の疾患の治療に影響する相乗効果をもたらし得る。

【0088】

核酸を用いる遺伝子発現の変更

KS RCCをコードする遺伝子の調節5', 3'や他の調節領域に相補的なアンチセンス分子(DNA、RNAまたはPNA)を設計して遺伝子発現を変更することができる。転写開始部位に関連して設計されたオリゴヌクレオチドが好ましい。同様に、阻害はポリメラーゼ、転写因子または調節分子の結合を阻害する三重らせん塩基対を用いて達成することができる(Gee ら In: Huber and Carr (1994) Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, 163-177ページ)。相補配列もまた、mRNAとリボソームの間に結合するのを阻止することによって翻訳を阻止するように設計することができる。一実施態様では、ライブラリまたは複数のcDNAまたはその断片を調節配列、翻訳されていない配列を特異結合するものを同定するためにスクリーニングすることが可能である。

30

【0089】

リボザイムは酵素性RNA分子であり、RNAの特異的切断を触媒するためにリボザイムを用いることもできる。リボザイム作用のメカニズムは、GUA、GUU、GUC等の部位でのヌクレオチド鎖切断に先立つ相補的標的RNAへのリボザイム分子の配列特異性ハイブリダイゼーションに関与している。一度それらの部位が同定すると、オリゴヌクレオチドを機能不全にするような2次構造の特徴に対して同じ配列のオリゴヌクレオチドを評価することが可能になる。候補標的の適合性の評価も、リボヌクレアーゼ保護アッセイを用いて相補的オリゴヌクレオチドとのハイブリダイゼーションをテストすることによって行うことができる。

40

【0090】

本発明の相補的な核酸及びリボザイムは、in vitroまたはin vivoで組換え発現によりあるいは固相フォスフォアミダイト化合物を使用して調製することが可能であ

50

る。さらに、分子の 5' 末端、3' 末端、あるいはその両方においてフランキング配列を追加したり、分子の主鎖内においてホスホジエステラーゼ結合ではなくホスホロチオネートまたは 2' - O - メチルを使用したりすることにより細胞内の安定性を高め、半減期を長くするために RNA 分子を修飾することができる。修飾は PNA の産出に固有のものであり、他の核酸分子に拡大することができる。非従来型塩基、例えばイノシン、クエオシン (queosine)、ワイブトシン (wybutosine) 等を加えること、あるいはアデニン、シチジン、グアニン、チミン、及びウリジンにアセチル - 、メチル - 、チオ - 基の修飾によって内在性エンドヌクレアーゼに対する分子の使用可能性が低下する。

【0091】

スクリーニング及び精製アッセイ

KSRCC をコードする cDNA を使用して、特異結合親和性のためにライブラリあるいは複数の分子または化合物をスクリーニングする。ライブラリは、生物系における cDNA の活性、複製、転写あるいは翻訳を調節する DNA 分子、RNA 分子、PNA、ペプチド、そして転写因子、エンハンサーまたはリプレッサーその他のリガンドなどのタンパク質であり得る。アッセイには、特異結合を可能にする条件下で分子のライブラリと cDNA またはその断片を結合すること、そして一本鎖または適切ならば二本鎖分子を特異結合する少なくとも 1 つの分子を同定する特異結合を検出することが含まれる。

【0092】

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。

【0093】

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。

【0094】

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哺乳動物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。

【0095】

好適な実施例では、KSRCC またはその部分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の基質に固定させる (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うことが可能である。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

10

20

30

40

50

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。

【0096】

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、タンパク質またはオリゴペプチドまたはその部分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。

10

【0097】

薬理学

医薬品成分は、所定、所望の目的を達成するために有効量の活性成分を含む物質である。有効投与量の決定は、当業者の能力の範囲内で行う。任意の化合物に対して、細胞培養アッセイ或いは動物モデルにおいて、先ず治療の有効投与量を推定することができる。動物モデルは好適な濃度範囲及び投与経路を達成するためにも用い得る。このような情報を用いて、次にヒトに対する有益な投与量及び投与経路を決定することができる。

20

【0098】

医学的に効果的な薬用量は、症状や容態を回復させるタンパク質またはインヒビターなどの活性処方成分の量に関連する。そのような薬剤の治療有効度及び毒性は、細胞培養または動物実験における標準的な薬剤手法によって、例えばED₅₀（集団の50%の医薬的有効量）またはLD₅₀（集団の50%の致死量）を測定するなどして決定することができる。毒性効果と薬用効果の間の投与量の比は、治療指数であり、LD₅₀/ED₅₀比として表すことができる。高い治療指数を示すような医薬品成分が望ましい。細胞培養アッセイ及び動物実験から得られたデータは、ヒトに用いるための投与量の範囲を調剤するのに用いられる。

30

【0099】

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る。

40

【0100】

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。

【0101】

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

50

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。

【0102】

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。

【0103】

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關して相当の証拠がある。

【0104】

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。

【0105】

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。

【0106】

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細胞がキャリアー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。

【0107】

ヒト胚盤胞由来のES細胞は、in vitro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in vitro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。

【0108】

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ト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

10

20

30

40

50

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。

【0109】

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。

【0110】

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ため、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。

【0111】

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、この哺乳動物タンパク質をコードするcDNAにその新技術を用い得る。

【0112】

（実施例）

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。実施例の目的のために、ヒト腎臓（KIDNNO T32）ライブラリの準備について記載する。

【0113】

1 cDNAライブラリ作製

腎臓ライブラリ作成用の組織は、頭蓋内出血および脳血管事故により死亡した49才の白人男性から得た。POLYTRONホモジナイザー（Brinkmann Instruments, Westbury NY）を使用して、凍結組織をTRIZOL試薬（1gの組織/10 mlのTRIZOLの割合；Life Technologies）でホモジナイズして溶解させる。氷上での短いインキュベートの後に、クロロホルムを加え（1:5 v/v）、混合液を遠心分離機にかけて相に分離した。上部の水様相を新しい試験管に移して、イソプロパノールを加えてRNAを沈殿させた。RNAをRNアーゼのない水で再懸濁し、DNアーゼで処理した。RNAを酸性フェノール-クロロホルムで再抽出し、酢酸ナトリウムとエタノールで再沈殿した。OLIGOTEX mRNA精製キット（QIAGEN, Valencia CA）を用いて、ポリ(A+) RNAを単離し

10

20

30

40

50

て、cDNAライブラリを作製するために使用した。

2 pINCYプラスミドの作製

プラスミドをEcoRI制限酵素(New England Biolabs, Beverly MA)と共にpSPORT1プラスミド(Life Technologies)を消化して、クレノウ酵素(New England Biolabs)と2'-デオキシヌクレオチド5'-3リン酸(dNTP)を用いて張出部(overhang)末端を占めるとにより作製した。プラスミドは自己連結反応して、細菌性宿主の大腸菌株JM109に形質転換した。

【0114】

バクテリア(pSPORT1ΔRI)により産出した仲介プラスミドはEcoRIでは消化されず、HindIII(New England Biolabs)で消化され、張出部(overhang)末端は再びクレノウとdNTPで占めた。リンカー配列をリン酸化して、5'平滑末端上に連結反応させ、EcoRIで消化してから、自己連結反応させた。JM109宿主細胞への形質転換後に、プラスミドを単離してEcoRIとの優先的消化性を検査したが、HindIIIとの優先的消化性を検査しなかった。この基準にかなう単一のコロニーを、pINCYプラスミドと名付けた。

【0115】

NotI、EcoRI制限酵素を用いて準備したライブラリのcDNAを組み込む能力に関してプラスミドを試験した後に、幾つかのクローンを配列化して、多量のプラスミドを準備するための中から約0.8 kbの挿入を含む単一クローンを選択した。ライブラリ作製で使用するために、NotI、EcoRIで消化した後に、アガロースゲル上でプラスミドを単離して、QIAQUICKカラム(Qiagen)を用いて精製した。

【0116】

3 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精製のためり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、カルベニシリン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。

【0117】

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0.25 x - 1.0 x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。

【0118】

4 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、OLIGOソフトウェア(Molecular Biology Insights)を用いて、約22個から約30

10

20

30

40

50

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。

【0119】

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。

【0120】

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200nmolの各プライマー、 Mg^{2+} と $(NH_4)_2SO_4$ と β -メルカプトエタノールを含む反応バッファー、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。

【0121】

各ウェルのDNA濃度は、1X TE及び0.5 μ lの希釈していないPCR産物に溶解した100 μ lのPICO GREEN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 μ lを1%アガロースミニゲル上で電気泳動法によって解析し、どの反応が配列の伸長に成功したかを決定した。

【0122】

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T4 DNA(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で一晩培養した。

【0123】

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2：94Cで15秒間、ステップ3：60Cで1分間、ス

10

20

30

40

50

ステップ 4 : 72°C 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°C で 5 分間、ステップ 7 : 4 で保管。上記したように P I C O G R E E N 量的試薬 (M o l e c u l a r P r o b e s) 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 (D M S O ; 1 : 2 , v / v) で希釈し、D Y E N A M I C エネルギー転写シークエンシングプライマー、及び D Y E N A M I C D I R E C T サイクルシークエンシングキット (A P B) または A B I P R I S M B I G D Y E ターミネーターサイクルシークエンシング反応キット (A p p l i e d B i o s y s t e m s) を用いてシークエンシングした。

【0124】

10

5 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、G e n B a n k、S w i s s P r o t、B L O C K S などのデータベースに問合せした。B L A S T か B L A S T 2 (前出の A l t s c h u l r ; 前出の A l t s c h u l) 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、S m i t h と S m i t h (1992 , P r o t e i n E n g i n e e r i n g 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 が記録されている) 。

20

【0125】

前出の K a r l i n 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 B L A S T 一致を統計的に評価して、ヌクレオチドに $10^{-2.5}$ 、ペプチドに $10^{-2.5}$ 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。B L A S T 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 B L A S T 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を約 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 % 一致まで設定した。

30

【0126】

無料で入手可能な配列比較アルゴリズムである B L A S T ソフトウェア一式 (N C B I , B e t h e s d a M D、<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「b l a s t n」、核酸分子またはアミノ酸分子を対で直接比較するために用いる B L A S T 2 など多種の配列分析プログラムが含まれる。B L A S T 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 : M a t r i x : B L O S U M 6 2 ; R e w a r d f o r m a t c h : 1 ; P e n a l t y f o r m i s m a t c h : - 2 ; O p e n G a p 5 及び E x t e n s i o n G a p : 2 p e n a l t i e s ; G a p x d r o p - o f f : 50 ; E x p e c t : 10 ; W o r d S i z e : 11 ; 及び F i l t e r : o n 同一性は配列全体またはより小さいその部分の長さにより測定する。B r e n n e r ら (1998 ; P r o c N a t l A c a d S c i 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 B L A S T 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。

40

【0127】

この出願の哺乳動物 cDNA をアセンブルしたコンセンサス配列あるいは L I F E S E Q G O L D データベースで見つけた鋳型と比較した。cDNA、伸長、全長およびショッ

50

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 P H R E D 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、A l u 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 b p であり、ジヌクレオチド反復、A l u 反復などの情報に乏しい配列と反復エレメントを " N s " あるいはマスクしたものに置換した。

【0128】

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。B L A S T と C R O S S M A T C H 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 B L A S T 質のスコアと少なく 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。P H R A P を用いて、各ピンの配列を構築した。D E E P P H R A P 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。

10

【0129】

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。S T I T C H E R / E X O N M A P P E R 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 B L A S T を用いて G b p r i などの G e n B a n k 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 - v a l u e (または確率値) $< 1 \times 10^{-8}$ を有すると定義した。鋳型を G E N P E P T に対するフレームシフト型 F A S T x にもかけた。相同一致を E - v a l u e $< 1 \times 10^{-8}$ 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 U S S N 09 / 276, 534 に記載されている。

20

【0130】

構築に続いて、鋳型を B L A S T、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 U S S N 08 / 812, 290、U S S N 08 / 811, 758、1997 年 10 月 9 日に提出した U S S N 08 / 947, 845、1998 年 3 月 4 日に提出した U S S N 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 H M M E R 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 P F A M 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。I n c y t e L i f e G o l d ヒト鋳型データベースを B L A S T 2 分析にかけ、クローン 3481942CB1 が 279978.9 鋳型の 5' 末端であることを発見した。更に分析して、3' 末端が関連する鋳型のクローン 4331016CT1であることを示した。S E Q I D N O : 2 を作製するために、これらの 2 つの鋳型を共にスプライシングした。M A C D N A S I S P R O ソフトウェア (Hitachi Software Engineering) と L A S E R G E N E ソフトウェア (DNA S T A R) を用いて c D N A を更に分析した。そして公共のデータベース、例えば G e n B a n k 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、S w i s s P r o t、B L O C K S、P R I N T S、P F A M そして P r o s i t e に問い合わせた。

30

40

【0131】

50

6 染色体マッピング

スタンフォード・ヒトゲノムセンター (S H G C)、ホワイトヘッド・ゲノム研究所 (W I G R)、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 c D N A 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 K S R C C をコードする c D N A 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。c M 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 D N A の 1 メガベース (M b) にほぼ等しい)。

【0132】

7 ハイブリダイゼーション技術と分析

10

基質上での c D N A の固定

下記の方法の 1 つによって c D N A を基質に適用する。ゲル電気泳動法により c D N A 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、c D N A 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 c D N A 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ベニシリン、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 L B 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% S D S、変性溶液 (1.5 M N a C l, 0.5 M N a O H)、中和溶液 (1.5 M N a C l, 1 M T r i s, p H 8.0)、2 x S S C (2 回) に各 10 分、連続的に膜を置く。次に膜を S T R A T A L I N K E R U V 架橋剤 (S t r a t a g e n e) で U V 照射する。

20

【0133】

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、c D N A を 30 サイクルの P C R で細菌ベクターから増幅する。P C R 増幅で 1 ~ 2 n g の初期量の核酸から 5 μ g より大きい最終量まで増大する。S E P H A C R Y L - 400 ビーズ (A P B) を用いて長さが 400 b p から約 5000 b p まで増幅した核酸を精製する。手であるいはドット/スロットプ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、U V 照射により固定する。U S P N 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 (C o r n i n g, A c t o n M A) を良く洗って 0.1% の S D S 中で超音波をかけ、4% フッ化水素酸 (V W R S c i e n t i f i c P r o d u c t s, W e s t C h e s t e r P A) 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 (S i g m a A l d r i c h) を用いてコーティングする、そして 110 のオブ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 S T R A T A L I N K E R U V 架橋剤 (S t r a t a g e n e) を用いてアレイを U V 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% S D S で洗浄し、蒸留水で 3 度洗浄する。リン酸緩衝生理食塩水 (P B S) (T r o p i x, I n c., B e d f o r d M A) 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% S D S 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。

30

40

【0134】

膜ハイブリダイゼーションのプロープ調製

配列表の c D N A 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 c D N A、m R N A またはゲノム D N A をスクリーニングする。c D N A を 45 μ l T E バッファの濃度 40 - 50 n g 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 c D N A を R E D I P R I M E 試験管 (A P B) を加え、青色が均一に分布するまで軽く混合して、

50

次に短時間遠心分離した。5 μ l の [32 P] d C T P をその試験管に加え、内容物を 37 で 10 分インキュベートする。5 μ l の 0.2 M E D T A を加えて標識化反応を停止する。そして P R O B E Q U A N T G - 50 m i c r o c o l u m n (A P B) を用いてプローブを組み込まれていないヌクレオチドから精製する。精製したプローブを 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。

【0135】

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ブ調製

サンプルから分離した m R N A に由来するハイブリダイゼーションのプローブを使用して、アレイベースのハイブリダイゼーションで配列表の c D N A をスクリーニングする。9 μ l T E バッファーで濃度 200 ng に m R N A を希釈し、5 μ l 5x バッファー、1 μ l 0.1 M D T T、3 μ l C y 3 または C y 5 標識化混合液、1 μ l R N アーゼ阻害因子、1 μ l 逆転写酵素、5 μ l 1x 酵母調節 m R N A を加え、G E M b r i g h t キット (I n c y t e G e n o m i c s) を用いてプローブを調製する。非コード酵母ゲノム DNA から *i n v i t r o* 転写により酵母調節 m R N A を合成する (W . L e i , 未発表)。定量用対照として、サンプル m R N A に 0.002 ng、0.02 ng、0.2 ng、2 ng での対照 m R N A 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。m R N A 示差発現パターンを調べるために、対照 m R N A 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 C H R O M A S P I N + T E 30 カラム (C l o n t e c h , P a l o A l t o C A) を利用してプローブを精製する。精製したプローブは、プローブを D E P C 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。プローブを 20 分間、20, 800 x g で遠心分離する。そしてペレットを 12 μ l の再懸濁バッファーで再懸濁し、5 分間 65 に加熱してから、十分に混合する。プローブを加熱して、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、プローブ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。

【0136】

膜系ハイブリダイゼーション

1% S a r k o s y l と 1x 高リン酸バッファ (0.5 M N a C l , 0.1 M N a ₂ H P O ₄ , 5 mM E D T A , p H 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ブを次に膜に加える。膜をプローブと共に 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 T r i s (p H 8.0) , 1% S a r k o s y l で 15 分間、25 で膜を洗浄し、さらに 1mM T r i s (p H 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、X O M A T - A R フィルム (E a s t m a n K o d a k , R o c h e s t e r N Y) 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。

【0137】

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

プローブを 5 分間 65 に加熱してから、5415 C マイクロ遠心分離で 5 分間、9400 rpm で遠心分離する (E p p e n d o r f S c i e n t i f i c , W e s t b u r y N Y) 。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 5x S S C 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ェンバーを、60 で約 6.5 時間インキュ

ベートする。アッセイを $1 \times \text{SSC}$, $0.1\% \text{ SDS}$ 中で、45 で10分間洗浄し、次に $0.1 \times \text{SSC}$ で、45 で10分間の洗浄を3回繰り返した後に、乾燥させた。

【0138】

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 (Shal on WO 95 / 35505)。

10

【0139】

ハイブリダイゼーション複合体は、Cy 3 の励起のためには 488 nm 、Cy 5 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、 $20 \mu\text{m}$ 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、Cy 3 では 565 nm 、Cy 5 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

【0140】

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

30

40

【0141】

8 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。ヒトおよびラット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

50

さになり、少なくとも積スコア 70 では一致は厳密となる。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。

【0142】

図 3 で示すように、電子ノーザン分析を積スコア 70 で実施した。L I F E S E Q データベースのすべての配列と c D N A 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

【0143】

9 相補的分子

c D N A に相補的な分子、約 5 (P N A) から 5 0 0 0 b p (c D N A 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。これらの分子を O L I G O ソフトウェア (M o l e c u l a r B i o l o g y I n s i g h t s) を使用して選択した。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 ' U T R 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 m R N A に結合するのを阻害する。

20

【0144】

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。

【0145】

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 (U S P N 4 , 7 3 6 , 8 6 6) 。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 c D N A 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

【0146】

10 K S R C C 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、哺乳動物タンパク質の発現と精製を達成する。p U B 6 / V 5 - H i s ベクター系 (I n v i t r o g e n , C a r l s b a d C A) を用いて、C H O 細胞の K S R C C を発現する。ベクターは選択可能 b s d 遺伝子、多数のクロニング部位、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 - V 5 抗体での抗体検出のための C - 末端 V 5 エピトープ、P R O B O N D 樹脂 (I n v i t r o g e n) 上での急速な精製のための C - 末端ポリヒスジン (6 x H i s) 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。

40

【0147】

S p o d o p t e r a f r u g i p e r d a (S f 9) 昆虫細胞を組換え A u t o g r a p h i c a c a l i f o r n i c a 核多角体病ウイルス (A c M N P V) で感染させる。相同組換えによって多角体遺伝子を哺乳動物 c D N A と置換する。多角体プロモーターにより c D N A の転写が行われる。上述の精製を可能にする 6 x h i s で、融合タンパ

50

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。

【0148】

11 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KSRCC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KSRCC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

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。

【0149】

12 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、pH2~3のバッファー、或いは高濃度の尿素またはチオシアン酸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。

【0150】

13 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。

【0151】

14 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、本発明の哺乳動物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菌を、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 1 x 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 - ブロモ - 4 - クロロ - 3 - インドリル β - 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 β - ガラクトシダーゼの発現を活性化する。

10

【0152】

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。ヒスチジンを含まない培地では育成せず、ヒスチジンを含む SD で育成するコロニーは、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 / - Ura で育成する。そして白色コロニーを単離して、増殖する。哺乳動物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。

20

【0153】

15 KSRCC アッセイ

¹²⁵I - 標識された KSRCC の存在下で候補リガンド分子を用いて KSRCC の AMP 結合活性をリガンド結合アッセイで決定する。KSRCC を ¹²⁵I ボルトンハンター試薬で標識する。(例えば Bolton A. E. および W. M. Hunter (1973) Biochem. J. 133: 529 - 539 を参照)。マルチウェルプレートのウェルに予め配列しておいた候補 AMP 分子を、標識した KSRCC と共にインキュベートし、洗浄して、標識した KSRCC 複合体を有する全てのウェルをアッセイする。様々な KSRCC 濃度で得られたデータを用いて、候補分子と結合した KSRCC の数量及び親和性、会合についての値を計算する。

30

【0154】

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。

40

【図面の簡単な説明】

【図1A】図1AはcDNA (SEQ ID NO: 2) によってコードされたKSRCC (SEQ ID NO: 1) を示す。アラインメントはMACDNASIS PROソフトウェア(日立ソフトウェアエンジニアリング, South San Francisco CA)を用いて作成した。

【図1B】図1BはcDNA (SEQ ID NO: 2) によってコードされたKSRCC (SEQ ID NO: 1) を示す。アラインメントはMACDNASIS PROソフトウェア(日立ソフトウェアエンジニアリング, South San Francisco CA)を用いて作成した。

50

【図1C】図1CはcDNA (SEQ ID NO: 2) によってコードされたKS RCC (SEQ ID NO: 1)を示す。アラインメントはMACDNASIS PROソフトウェア (日立ソフトウェアエンジニアリング, South San Francisco CA)を用いて作成した。

【図1D】図1DはcDNA (SEQ ID NO: 2) によってコードされたKS RCC (SEQ ID NO: 1)を示す。アラインメントはMACDNASIS PROソフトウェア (日立ソフトウェアエンジニアリング, South San Francisco CA)を用いて作成した。

【図1E】図1EはcDNA (SEQ ID NO: 2) によってコードされたKS RCC (SEQ ID NO: 1)を示す。アラインメントはMACDNASIS PROソフトウェア (日立ソフトウェアエンジニアリング, South San Francisco CA)を用いて作成した。 10

【図1F】図1FはcDNA (SEQ ID NO: 2) によってコードされたKS RCC (SEQ ID NO: 1)を示す。アラインメントはMACDNASIS PROソフトウェア (日立ソフトウェアエンジニアリング, South San Francisco CA)を用いて作成した。

【図2A】図2Aは、KS RCC (SEQ ID NO: 1)、ドブネズミKS (g3127193)、ヒトKS (g3219339)、SEQ ID NO: 1, 15 - 16のドメインの保存した化学的及び構造的類似性が示す。アラインメントは、LASERGENEソフトウェア (DNASTAR, Madison WI)のMEGALIGN programを使用して作製された。 20

【図2B】図2B、KS RCC (SEQ ID NO: 1)、ドブネズミKS (g3127193)、ヒトKS (g3219339)、SEQ ID NO: 1, 15 - 16のドメインの保存した化学的及び構造的類似性が示す。アラインメントは、LASERGENEソフトウェア (DNASTAR, Madison WI)のMEGALIGN programを使用して作製された。

【図3】図3は、LIFESEQ Gold database (Incyte Genomics, Palo Alto CA)を使用して作製したKS RCCのノーザン分析を示す。列1は組織のカテゴリ、列2は組織のカテゴリのクローニングの数、列3は少なくとも1つの転写物が見られるライブラリの数、列4は転写物の絶対的存在量、列5は転写物の存在量%を示す。 30

【図4A】図4Aは、KS RCC、ラットg3127193とヒトg3219339の親水性プロットと抗原指数を示す。

【図4B】図4Bは、KS RCC、ラットg3127193とヒトg3219339の親水性プロットと抗原指数を示す。

【 図 1 A 】

[illegible]

1A

【 図 1 B 】

[illegible]

紅 樹

【 図 1 C 】

186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

2.1

【 叉 1 D 】

[illegible]

CT

【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
28 February 2002 (28.02.2002)

PCT

(10) International Publication Number
WO 02/16595 A2

(51) International Patent Classification: C12N 15/00

(21) International Application Number: PCT/US01/26317

(22) International Filing Date: 23 August 2001 (23.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data: 09/645,961 24 August 2000 (24.08.2000) US

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

(72) Inventors; and
(75) Inventors/Applicants (for US only): WALKER, Michael G., [CA/US], 1050 Boregas Avenue, #80, Sunnyvale, CA 94089 (US); KRASNOW, Randi, E., [US/US], 817 Santa Fe Avenue, Stanford, CA 94305 (US).

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

(81) Designated States (national): AR, AU, AT, AM, AE, 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, GR, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW); Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM); European patent (AT, BE, CH, CY, DE, DK, ES, FR, GB, GR, IE, IT, LU, NL, PT, SE, TR); OAPI patent (BH, BI, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published: — without international search report and to be republished upon receipt of that report

[Continued on next page]

(54) Title: KIDNEY-SPECIFIC PROTEIN

6' GGT GCT TCT CTT CCA AGG CTG TAG GAG TTC TGG AGC TGC TGG CTG GAG AGG AGG UUU
 ...
 64 73 82 91 100 109
 GAG GAG GGT CTT AGA AAG ACA TCC TGA GAG GAG TTC GCA GCA CCA CCA AAT AAT
 ...
 118 127 136 145 154 163
 CAT TGG CTG CGA AAA GTT GAG GGA CTT TAC AGC TGG TGG GGT ACT CAG AAT TCC
 ...
 172 181 190 199 208 217
 AAC CCG AGT CTT TAC ATT AAT AGT AAG CAA CTG GTG TCC CTG CAG CAG GAT CAG
 ...
 226 235 244 253 262 271
 GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG
 ...
 280 289 298 307 316 325
 GAG ATG GAG AAG OCT GGC AAG GCA CTC CCA AGC CCA GGC GGC GGC GGC GGC GGC GGC
 ...
 334 343 352 361 370 379
 GAG AAG GGC AAG GAG TTA AAT TGG TGC TGC TGC TGC TGC TGC TGC TGC TGC TGC TGC
 ...

(57) Abstract: The invention provides a mammalian cDNA which encodes a kidney-specific 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 kidney disorders.

WO 02/16595 A2

WO 02/16595 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/16595

PCT/US01/26317

KIDNEY-SPECIFIC PROTEIN**TECHNICAL FIELD**

5 This invention relates to a mammalian cDNA which encodes a kidney-specific protein which is diagnostic of renal cell carcinoma (KSRCC) and to the use of the cDNA and the encoded protein in the diagnosis and treatment of kidney disorders.

BACKGROUND OF THE INVENTION

10 Phylogenetic relationships among organisms have been demonstrated many times, and studies 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
15 accelerate the investigation of human sequences and allow the development of model systems for testing diagnostic and therapeutic agents for human conditions, diseases, and disorders.

The human kidneys are two bean-shaped organs on either side of the backbone. The kidneys filter waste from the blood, form urine, and regulate the water and electrolyte content of the body. They reabsorb and retain proteins, glucose, amino acids, and bicarbonate and inorganic
20 phosphate, and release hormones which interact to maintain intravascular volume. They contribute to maintenance of the body's acid/base balance and of homeostasis by regulating blood pH electrolyte levels and blood pressure.

Renal cell carcinoma (RCC), originating within the proximal renal tubular epithelium, is the most common type of kidney cancer accounting for approximately 3% of all adult cancers. There
25 are four main types of renal cell carcinoma, with the majority being clear cell type, and mixed granular and clear cell type. RCC is curable only in its very early stages through surgery. It is resistant to chemotherapy and relatively resistant to radiotherapy. Risk factors for development of RCC include cigarette smoking and development of acquired cystic kidney disease.

The discovery of a mammalian cDNA encoding KSRCC satisfies a need in the art by
30 providing compositions which are useful in the diagnosis and treatment of kidney disorders, particularly, renal cell carcinoma.

SUMMARY OF THE INVENTION

The invention is based on the discovery of a mammalian cDNA which encodes a mammalian kidney specific-protein which is useful in the diagnosis and treatment of kidney
35 disorders including acquired cystic kidney disease and, particularly, renal cell carcinoma.

The invention provides an isolated mammalian cDNA or a fragment thereof encoding a

WO 02/16595

PCT/US01/26317

mammalian protein or a portion thereof selected from the group consisting of an amino acid sequence of SEQ ID NO:1, a variant having 80% identity to the amino acid sequence of SEQ ID NO:1, an antigenic epitope of SEQ ID NO:1, an oligopeptide of SEQ ID NO:1, and a biologically active portion of SEQ ID NO:1. The invention also provides an isolated mammalian cDNA or the complement thereof selected from the group consisting of a nucleic acid sequence of SEQ ID NO:2, a variant having 83% identity to the nucleic acid sequence of SEQ ID NO:2, a fragment of SEQ ID NOs:3-14, an oligonucleotide of SEQ ID NOs:2-14. The invention additionally provides a composition, a substrate, and a probe comprising the cDNA, or the complement of the cDNA, encoding KSRCC. The invention further provides a vector containing the cDNA, a host cell containing the vector and a method for using the cDNA to make KSRCC. The invention still further provides a transgenic cell line or organism comprising the vector containing the cDNA encoding KSRCC. The invention additionally provides a mammalian fragment or the complement thereof selected from the group consisting of SEQ ID NOs:11-14. In one aspect, the invention provides a substrate containing at least one of these fragments. In a second aspect, the invention provides a probe comprising the fragment 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 renal cell carcinoma. In another aspect, the cDNA or a fragment or a complement thereof may comprise an element on an array.

The invention additionally provides a method for using a cDNA or a fragment or a complement 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 mammalian protein or a portion thereof selected from the group consisting of an amino acid sequence of SEQ ID NO:1, a variant having 80% identity to the

WO 02/16595

PCT/US01/26317

amino acid sequence of SEQ ID NO:1, an antigenic epitope of SEQ ID NO:1, an oligopeptide 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 or a portion thereof in conjunction with a pharmaceutical carrier. The invention further provides a method of using the KSRCC to treat a subject with renal cell carcinoma 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 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 renal cell carcinoma.

The invention provides a method of using a mammalian 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 standards, wherein the presence or quantity of antibody is diagnostic of renal cell carcinoma.

The invention also provides a method of using a mammalian protein to prepare and purify antibodies comprising immunizing an 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 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 renal cell carcinoma. The invention also provides a method of using an antibody to diagnose renal cell carcinoma comprising combining the antibody comparing the quantity of bound antibody to known standards, thereby establishing the presence of renal cell carcinoma. The invention further provides a method of using an antibody to treat renal cell carcinoma comprising administering to a patient in need of such treatment a pharmaceutical composition comprising the purified antibody.

The invention provides a method for inserting a 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:2-14, transforming the

WO 02/16595 PCT/US01/26317
 vector into an embryonic stem cell, selecting a transformed embryonic stem, microinjecting the transformed 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
 5 mammal to produce a homozygous, mammalian model system.

BRIEF DESCRIPTION OF THE FIGURES AND TABLE

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

10 Figures 2A and 2B demonstrate the conserved chemical and structural similarities among the domains of KSRCC (SEQ ID NO:1), Rattus norvegicus KS (g3127193), and Homo sapiens KS (g3219339). SEQ ID NOs: 1, and 15-16, respectively. The alignment was produced using the MEGALIGN program of LASERGENE software (DNASTAR, Madison WI).

Figure 3 shows the northern analysis for KSRCC produced using the LIFESEQ Gold
 15 database (Incyte Genomics, Palo Alto CA). The first column presents the tissue categories; the second column, the number of clones in the tissue category; the third column, the number of libraries in which at least one transcript was found; the fourth column, absolute abundance of the transcript; and the fifth column, percent abundance of the transcript.

Figure 4 shows the hydrophobicity plots and antigenic indices for KSRCC, rat g3127193 and
 20 human g3219339.

DESCRIPTION OF THE INVENTION

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
 25 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 to "a host cell" includes a plurality of such host cells known to those skilled in the art.

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
 30 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

35 "KSRCC" refers to a substantially purified protein obtained from any mammalian species.

WO 02/16595

PCT/US01/26317

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 sequence, and the other, a cDNA of diagnostic

5 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 a cDNA and a 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
10 mRNA under conditions of high stringency.

"cDNA" refers to an isolated polynucleotide, nucleic acid molecule, or any fragment or complement thereof. It may have originated recombinantly or synthetically, be double-stranded or single-stranded, represent coding and/or noncoding sequence, an exon with or without an intron from a genomic DNA molecule.

15 The phrase "cDNA encoding a protein" refers to a nucleic acid 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; Altschul (1993) *J Mol Evol* 36: 290-300; Altschul et al. (1990) *J Mol Biol* 215:403-410) which provides identity within the conserved region.

20 "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
25 occurring molecules but may confer advantages such as longer lifespan or enhanced activity.

"Differential expression" refers to an increased, upregulated or present, or decreased, downregulated or absent, gene expression as detected by the absence, presence, 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 KSRCC are
30 differentially expressed.

"Fragment" refers to a chain of consecutive nucleotides from about 200 to about 700 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. Nucleic acids and their ligands identified in this manner are useful as therapeutics to regulate replication, transcription or

35 translation.

WO 02/16595

PCT/US01/26317

"Guilt by association" (GBA) is a method for identifying cDNAs or proteins that are associated with a specific disease, regulatory pathway, subcellular compartment, cell type, tissue type, or species. In particular, the method identifies cDNAs useful in diagnosis, prognosis, treatment, and evaluation of therapies for kidney disorders.

5 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'. The degree of complementarity and the use of nucleotide analogs affect the efficiency and stringency of hybridization reactions.

"Ligand" refers to any agent, molecule, or compound which will bind specifically to a 10 complementary site on a cDNA molecule or polynucleotide, or to an epitope or a protein. Such ligands stabilize or modulate the activity of polynucleotides or proteins and may be composed of inorganic or organic substances including nucleic acids, proteins, carbohydrates, fats, and lipids.

"Oligonucleotide" refers to 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 15 replication, transcription or translation. Substantially equivalent terms are amplicon, 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, 20 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 25 reporter molecules for use in hybridization reactions including Southern, northern, in situ, dot blot, array, and like technologies or in screening assays.

"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

30 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 35 naturally associated.

WO 02/16595

PCT/US01/26317

"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
5 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, the hydrogen bonding along the backbone between two single stranded nucleic acids, or the binding between an
10 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
15 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.

"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
20 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)
25 refers to a 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

The invention is based on the discovery of a cDNA which encodes KSRCC and on the use
30 of the cDNA, or fragments thereof, and protein, or portions thereof, directly or as compositions in the characterization, diagnosis, and treatment of kidney disorders.

Nucleic acids encoding the KSRCC of the present invention were first identified (in Incyte Clone 3481942CB1 from kidney tissue cDNA library, KIDNNOT31), as kidney-specific through GBA analysis of sequences co-regulated in kidney pathways. Eleven known kidney
35 disease-associated genes were selected to identify novel genes that are closely associated with

WO 02/16595 PCT/US01/26317
 diseases of the kidney. These known genes were uromodulin, NKCC2, NCCT, aldolase B, ROMK1, ATP1G1, PDZK1, NPT-1, calbindin, kininogen, and CIC-Kb. The degree of association was measured by probability values using a cutoff p value less than 0.00001. The sequences were further examined to ensure that the genes that passed the probability test had strong association with
 5 known kidney disease-associated genes. Nine novel genes showed strong association with known kidney disease-associated genes from a total of 41,419. Each of the nine novel genes was coexpressed with at least one of the 11 known genes with a p-value of less than $10e^{-05}$. KSRCC cDNA is 2054 nucleic acids in length. SEQ ID NO:3 (Incyte Clone 3481942CB1) is 521 nucleotides in length and has 82% identity using BLAST2 to the nucleic acid sequence of KS, a rat
 10 kidney-specific protein (g3127192).

The cDNA for SEQ ID NO:2 was derived from the following overlapping and/or extended nucleic acid sequences (SEQ ID NO:3-10): Incyte Clones 3481942CB1, 5519427H1 (L1VRDIR01), 5390984F8 (KIDNNOT32), 4332230T9 (KIDNNOT32), 760835R1 (BRAITUT02), 762540R1 (BRAITUT02), 2099420R6 (BRAITUT02), and 7695905J1 (LGcompseqsUN2000new).

15 Figure 3 shows expression of the transcript in kidney tissue, particularly in tissues from patients with renal cell carcinoma. Therefore, the cDNA is useful in assays to diagnose renal cell carcinoma. A fragment of SEQ ID NO:2 from nucleotide 1 to nucleotide 106 is useful as a hybridization probe. An oligonucleotide from 54 to 69 or 80 to 95 is useful as a diagnostic to distinguish the transcript encoding KSRCC from other kidney-specific proteins.

20 In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1 as shown in Figures 1A through 1F. KSRCC is 577 amino acids in length and has 77% identity to the amino acid sequence of the rat kidney-specific protein (g3127193) using BLAST2 analysis. BLIMPS analysis shows that SEQ ID NO:1 has a 16 amino acid segment that matches the AMP-binding domains of g3127193. BLOCKS analysis indicates
 25 that the regions of KSRCC from Q213 to T224 and S225 to H233 are similar to AMP-binding sites. Pfam analysis indicates that the region of KSRCC from N82 to V493 is similar to an AMP-binding site.

As shown in Figures 2A and 2B, KSRCC has chemical and structural homology with kidney-specific protein from rat (g3127193:SEQ ID NO:15) and a human homolog of the rat protein
 30 (g3219339:SEQ ID NO:16). In particular, KSRCC and the rat kidney-specific protein share 77% sequence identity over proteins of nearly identical length (577 AA residues and 572 AA residues, respectively). The human homolog of the rat protein (g3219339) is a shorter protein (207 AA residues), but is 97% identical to KSRCC over this entire length.

As shown in Figures 4A and 4B, the three proteins have similar hydrophobicity plots and
 35 antigenic indices.

WO 02/16595

PCT/US01/26317

Mammalian variants of the cDNA encoding KSRCC were identified using BLAST2 with default parameters and the ZOOSEQ databases (Incyte Genomics). Mammalian variants of the cDNA encoding the KSRCC include 701648693H1 (RALITXT40), 212451_Rn.1 (template), 202264_Rn.1 (template), and 701836346T1 (RAKITXT11); SEQ ID NOs:11-14 of the Sequence Listing, respectively. These cDNAs are particularly useful for producing transgenic cell lines or organisms which model human kidney disorders and upon which potential therapeutic treatments for such kidney disorders may be tested. Table 1 below shows the percent identity between the cDNA encoding KSRCC and its mammalian variants, SEQ ID NOs:11-14. The first column shows the SEQ ID for the human cDNA; the second column, the SEQ IDvar for variant cDNAs; the third column, the clone number for the variant cDNAs; the fourth column, the percent identity to the human cDNA; and the fifth column, the alignment of the variant cDNA to the human cDNA.

SEQ ID _h	SEQ ID _{var}	Clone _{var}	Identity	N ₁₄ Alignment
2	11	701648693H1	83%	407-513
2	12	212451_Rn.1	84%	709-986
2	13	202264_Rn.1	83%	1067-1764
2	14	701836346T1	84%	1655-1812

These cDNAs are particularly useful for producing transgenic cell lines or organisms which model human kidney disorders and upon which potential therapeutic treatments for such kidney disorders may be tested.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of cDNA encoding KSRCC, some bearing minimal similarity to the cDNAs of 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 HSTK, and all such variations are to be considered as being specifically disclosed.

The cDNA and fragments thereof (SEQ ID NOs:2-14) 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 renal cell carcinoma 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

GBA Analysis

GBA identifies cDNAs that are expressed in a plurality of cDNA libraries. The cDNAs

WO 02/16595

PCT/US01/26317

include genes of known or unknown function which are expressed in a specific disease process, subcellular compartment, cell type, tissue type, or species. The expression patterns of genes with known function are compared with those of cDNAs with unknown function to determine whether a specified co-expression probability threshold is met. Through this comparison, a subset of the cDNAs having a high co-expression probability with the known genes can be identified. The high co-expression probability correlates with a particular coexpression probability threshold which is preferably less than 0.001 and more preferably less than 0.00001.

The cDNAs originate from cDNA libraries derived from a variety of sources including, but not limited to, eukaryotes such as human, mouse, rat, dog, monkey, plant, and yeast; prokaryotes such as bacteria; and viruses. These cDNAs can also be selected from a variety of sequence types including, but not limited to, expressed sequence tags (ESTs), assembled polynucleotides, full length gene coding regions, promoters, introns, enhancers, 5' untranslated regions, and 3' untranslated regions. To have statistically significant analytical results, the cDNAs need to be expressed in at least five cDNA libraries.

The cDNA libraries used in the co-expression analysis of the present invention can be obtained from any cell or cell line, tissue, or organ and may be from adrenal gland, biliary tract, bladder, blood cells, blood vessels, bone marrow, brain, bronchus, cartilage, chromaffin system, colon, connective tissue, cultured cells, embryonic stem cells, endocrine glands, epithelium, esophagus, fetus, ganglia, heart, hypothalamus, immune system, intestine, islets of Langerhans, kidney, larynx, liver, lung, lymph, muscles, neurons, ovary, pancreas, penis, peripheral nervous system, phagocytes, pituitary, placenta, pleurus, prostate, salivary glands, seminal vesicles, skeleton, spleen, stomach, testis, thymus, tongue, ureter, uterus, and the like. The number of cDNA libraries selected can range from as few as 500 to greater than 10,000.

Known kidney specific genes can be selected based on the use of the genes as diagnostic or prognostic markers or as therapeutic targets for diseases associated with kidney disorders.

The method for identifying cDNAs that exhibit a statistically significant co-expression pattern with known kidney specific genes is as follows. First, the presence or absence of a gene sequence in a cDNA library is defined: a gene is present in a cDNA library when at least one cDNA fragment corresponding to that gene is detected in a cDNA sample taken from the library, and a gene is absent from a library when no corresponding cDNA fragment is detected in the sample.

Second, the significance of gene co-expression is evaluated using a probability method to measure a due-to-chance probability of the co-expression. The probability method can be the Fisher exact test, the chi-squared test, or the kappa test. These tests and examples of their applications are well known in the art and can be found in standard statistics texts (Agresti (1990) Categorical Data Analysis, John Wiley & Sons, New York NY; Rice (1988) Mathematical Statistics and Data

WO 02/16595

PCT/US01/26317

Analysis, Duxbury Press, Pacific Grove CA). A Bonferroni correction (Rice, *supra*, p. 384) can also be applied in combination with one of the probability methods for correcting statistical results of one gene versus multiple other genes. In a preferred embodiment, the due-to-chance probability is measured by a Fisher exact test, and the threshold of the due-to-chance probability is set

5 preferably to less than 0.001, more preferably to less than 0.00001.

To determine whether two genes, A and B, have similar co-expression patterns, occurrence data vectors can be generated as illustrated in the table below. The presence of a gene occurring at least once in a library is indicated by a one, and its absence from the library, by a zero.

10		Library 1	Library 2	Library 3	...	Library N
	Gene A	1	1	0	...	0
	Gene B	1	0	1	...	0

For a given pair of genes, the co-occurrence data can be summarized in a 2 x 2 contingency table.

15		Gene A Present	Gene A Absent	Total
	Gene B Present	8	2	10
	Gene B Absent	2	18	20
	Total	10	20	30

20 The contingency table shows the co-occurrence data for gene A and gene B in a total of 30 libraries. Both gene A and gene B occur 10 times in the libraries, and the table summarizes and presents: 1) the number of times gene A and B are both present in a library; 2) the number of times gene A and B are both absent in a library; 3) the number of times gene A is present, and gene B is absent; and 4) the number of times gene B is present, and gene A is absent. The upper left entry is

25 the number of times the two genes co-occur in a library, and the middle right entry is the number of times neither gene occurs in a library. The off diagonal entries are the number of times one gene occurs, and the other does not. Both A and B are present eight times and absent 18 times. Gene A is present, and gene B is absent, two times; and gene B is present, and gene A is absent, two times. The probability ("p-value") that the above association occurs due to chance as calculated using a

30 Fisher exact test is 0.0003. Associations are generally considered significant if a p-value is less than 0.01 (Agresti, *supra*; Rice, *supra*).

This method of estimating the probability for co-expression of two genes makes several assumptions. The method assumes that the libraries are independent and are identically sampled. However, in practical situations, the selected cDNA libraries are not entirely independent, because

35 more than one library may be obtained from a single subject or tissue. Nor are they entirely

WO 02/16595

PCT/US01/26317

identically sampled, because different numbers of cDNAs may be sequenced from each library. The number of cDNAs sequenced typically ranges from 5,000 to 10,000 cDNAs per library. In addition, because a Fisher exact co-expression probability is calculated for each gene versus every other assembled gene that occur in at least five libraries, a Bonferroni correction for multiple

5 statistical tests is used.

cDNA libraries

In a particular embodiment disclosed herein, mRNA was 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 clones listed above were isolated from mammalian cDNA libraries.

10 Three library preparations representative of the invention are described in the EXAMPLES below. The consensus sequences were chemically and/or electronically assembled from fragments including Incyte clones 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). Clones, extension and/or shotgun sequences are

15 electronically assembled into clusters and/or master clusters.

Sequencing

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

20 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 MICROLAB 2200 system (Hamilton, Reno NV) and the DNA ENGINE thermal cycler (MJ Research, Watertown MA). Machines commonly used for sequencing include the ABI

25 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 in the art and described in Ausubel *et al.* (1997; Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7) and in Meyers (1995; Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

30 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,

35 heat-labile DNA polymerases, and primers chosen from representative regions flanking the cDNAs

WO 02/16595 PCT/US01/26317
 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
 5 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
 10 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
 20 KSRCC, 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:2-14. 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
 25 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, and temperature. In particular, stringency can be increased by reducing the concentration of salt or raising the hybridization temperature. In solutions used for some membrane
 30 based hybridizations, addition of an organic solvent such as formamide allows the reaction to occur at a lower 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
 35 45C (medium stringency) or 68C (high stringency). At high stringency, hybridization complexes

WO 02/16595 PCT/US01/26317
 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 other 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.

Arrays may be prepared and analyzed using methods known in the art. Oligonucleotides may be used as either probes or targets in an array. The array can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and single nucleotide polymorphisms. Such information 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; Baldeschwiler *et al.* (1995) PCT application WO95/251116; Shalon *et al.* (1995) PCT application WO95/35505; 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: 1) a particular chromosome, 2) a specific region of a chromosome, or 3) an artificial chromosome construction such as human artificial chromosome (HAC), yeast artificial chromosome (YAC), bacterial artificial chromosome (BAC), bacterial P1 construction, or single chromosome cDNA libraries.

Expression

Any one of a multitude of cDNAs encoding KSRCC 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 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,

WO 02/16595

PCT/US01/26317

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).

- 5 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.

- 10 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.

- 15 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
20 cells which successfully express the introduced sequences. Resistant clones identified either by survival on selective media or by the expression of visible markers, such as anthocyanins, green fluorescent protein (GFP), β glucuronidase, luciferase and the like, may be propagated using culture techniques. Visible markers are also used to quantify the amount of protein expressed by the introduced genes. Verification that the host cell contains the desired mammalian cDNA is based on
25 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
30 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.

Recovery of Proteins from Cell Culture

- Heterologous moieties engineered into a vector for ease of purification include glutathione
35 S-transferase (GST), 6xHis, FLAG, MYC, and the like. GST and 6xHis are purified using

WO 02/16595

PCT/US01/26317

commercially available affinity matrices such as immobilized glutathione and metal-chelate resins, respectively. 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.

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 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 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).

Preparation and Screening of Antibodies

Various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with KSRCC 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 response. The oligopeptide, peptide, or portion of protein used to induce antibodies should consist of 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

WO 02/16595

PCT/US01/26317

antibodies to the chimeric molecule.

Monoclonal antibodies may be prepared using any technique which provides for the 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

5 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* 6:109-120.)

Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce epitope specific single chain antibodies.

10 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 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, 15 e.g., Huse *et al.* (1989) *Science* 246:1275-1281.)

The KSRCC 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 protocols for competitive binding or immunoassays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays 20 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) *Immunochemical Protocols*, Humana Press, Totowa NJ).

Labeling of Molecules for Assay

25 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 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 30 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 BODIPY or FITC (Molecular Probes, Eugene OR).

DIAGNOSTICS

The cDNAs, fragments, oligonucleotides, complementary RNA and DNA molecules, and 35 PNAs and may be used to detect and quantify differential gene expression, absence/presence vs.

WO 02/16595

PCT/US01/26317

excess, expression of mRNAs or to monitor mRNA levels during therapeutic intervention. Similarly antibodies which specifically bind KSRCC may be used to quantitate the protein. Kidney disorders associated with differential expression include acquired cystic kidney disease and, in particular, renal cell carcinoma. 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 substantially 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 and in clinical trial 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.

30 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

WO 02/16595 PCT/US01/26317
 employed. (See, e.g., Coligan *et al.* (1997) *Current Protocols in Immunology*, Wiley-Interscience, New York NY; and Pound, *supra*.)

THERAPEUTICS

- Chemical and structural similarity, in the context of the AMP-binding domain, exists
 5 between regions of KSRCC (SEQ ID NO:1) and the kidney-specific proteins from rat and human
 (g312719, and g3219339, respectively) as shown in Figures 2A and 2B. In addition, differential
 expression is highly associated with tissues and with renal cell carcinoma as shown in Figure 3.
 KSRCC clearly plays a role in acquired cystic kidney disease and, in particular, renal cell
 carcinoma.
- 10 In the treatment of conditions associated with increased expression of the protein such as
 renal cell carcinoma, 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
 15 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.
- Any of the cDNAs, complementary molecules, or fragments thereof, proteins or portions
 20 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.
- 25 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
 KSRCC. Oligonucleotides designed with reference to the transcription initiation site are preferred.
 Similarly, inhibition can be achieved using triple helix base-pairing which inhibits the binding of
 30 polymerases, transcription 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 or fragments thereof may
 be screened to identify those which specifically bind a regulatory, nontranslated sequence.
- 35 Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage

WO 02/16595

PCT/US01/26317

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 evaluated for secondary structural features which would render the

5 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. 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 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 KSRCC 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, peptides, proteins such as transcription factors, enhancers, repressors, and other ligands which regulate the activity, replication, transcription, or translation of the cDNA in the biological system. The assay involves combining the cDNA or a fragment thereof with the 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, if appropriate, double stranded molecule.

25 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 initially determined by gel shift assay and may be later confirmed by recovering and raising antibodies against that molecule or compound. When these antibodies are added into the assay, they cause a supershift in the gel-retardation assay.

30 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 reacted with cyanogen bromide groups on a polymeric resin or gel. Then a sample is

WO 02/16595 PCT/US01/26317
 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 sample. A method for using a mammalian 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, KSRCC or a portion thereof may be used to screen a plurality of molecules or 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 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 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 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 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 or oligopeptide or portion thereof. Molecules or compounds identified by screening may be used in a mammalian model system to evaluate their toxicity, diagnostic, or therapeutic potential.

Pharmacology

Pharmaceutical compositions are those substances wherein the active ingredients are 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

WO 02/16595
administration in humans.

PCT/US01/26317

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.,
5 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.

10 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
15 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.

20 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
25 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
30 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 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.

35 Acute toxicity tests are based on a single administration of an agent to the subject to

WO 02/16595 PCT/US01/26317
 determine the symptomatology or lethality of the agent. Three experiments are conducted: 1) an initial 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 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 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 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 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 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.

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

WO 02/16595
Knockout Analysis

PCT/US01/26317

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 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.

10 Knockin Analysis

ES cells can be used to create knockin humanized animals (pigs) or transgenic animal models (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 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. 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 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 mammalian protein may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on 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

WO 02/16595

PCT/US01/26317

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention. For purposes of example, preparation of the human kidney (KIDNOT32) library will be described.

I cDNA Library Construction

5 The tissue used for kidney library construction was obtained from a 49-year-old Caucasian male who died from an intracranial hemorrhage and cerebrovascular accident. The frozen tissue was homogenized and lysed in TRIZOL reagent (1 g tissue/10 ml TRIZOL; Life Technologies) using a POLYTRON homogenizer (Brinkmann Instruments, Westbury NY). After brief incubation on ice, chloroform was added (1:5 v/v), and the mixture was centrifuged to
10 separate the phases. The upper aqueous phase was removed to a fresh tube, and isopropanol was added to precipitate RNA. The RNA was resuspended in RNase-free water and treated with DNase. The RNA was re-extracted with acid phenol-chloroform and reprecipitated with sodium acetate and ethanol. Poly(A+) RNA was isolated using the OLIGOTEX mRNA purification kit (QIAGEN, Valencia CA) and used to construct the cDNA library. .

15 II Construction of pINCY Plasmid

The plasmid was constructed by digesting the pSPORT1 plasmid (Life Technologies) with EcoRI restriction enzyme (New England Biolabs, Beverly MA) and filling the overhanging ends using Klenow enzyme (New England Biolabs) and 2'-deoxynucleotide 5'-triphosphates (dNTPs). The plasmid was self-ligated and transformed into the bacterial host, *E. coli* strain JM109.

20 An intermediate plasmid produced by the bacteria (pSPORT 1-ARI) showed no digestion with EcoRI and was digested with Hind III (New England Biolabs) and the overhanging ends were again filled in with Klenow and dNTPs. A linker sequence was phosphorylated, ligated onto the 5' blunt end, digested with EcoRI, and self-ligated. Following transformation into JM109 host cells, plasmids were isolated and tested for preferential digestibility with EcoRI, but not with Hind III. A
25 single colony that met this criteria was designated pINCY plasmid.

After testing the plasmid for its ability to incorporate cDNAs from a library prepared using NotI and EcoRI restriction enzymes, several clones were sequenced; and a single clone containing an insert of approximately 0.8 kb was selected from which to prepare a large quantity of the plasmid. After digestion with NotI and EcoRI, the plasmid was isolated on an agarose gel and
30 purified using a QIAQUICK column (Qiagen) for use in library construction.

III 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). This kit consists of a 96-well block with reagents for 960 purifications. The recommended protocol was
35 employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile

WO 02/16595

PCT/US01/26317

TERRIFIC BROTH (BD Biosciences, Sparks MD) 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
5 for storage at 4C.

The cDNAs were prepared for sequencing using the MICROLAB 2200 system (Hamilton) in 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
10 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 APE.

IV Extension of cDNA Sequences

The cDNAs were extended using the cDNA clone and oligonucleotide primers. One primer
15 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 OLIGO software (Molecular Biology Insights), 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.

20 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.

25 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
30 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
35 min; Step 7: storage at 4C.

WO 02/16595

PCT/US01/26317

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 mini-gel to determine which reactions were successful in extending the sequence.

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 AGARASE 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.

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 quantitative 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).

V Homology Searching of cDNA Clones and Their Deduced Proteins

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 BLAST 2 (Altschul et al. *supra*; Altschul, *supra*) 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. 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

WO 02/16595

PCT/US01/26317

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 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 the laboratory, the electronic stringency for an exact match was set at 70, and the conservative lower limit for an exact match was set at approximately 40 (with 1-2% error due to uncalled bases).

The BLAST software suite, freely available sequence comparison algorithms (NCBI, Bethesda MD; <http://www.ncbi.nlm.nih.gov/blast/>), includes various sequence analysis programs including "blastn" that is used to align nucleic acid molecules and BLAST 2 that is used for direct pairwise comparison of either nucleic or amino acid molecules. 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 or some smaller portion thereof. Brenner *et al.* (1998; Proc Natl Acad Sci 95:6073-6078, incorporated herein by reference) analyzed the 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 mammalian cDNAs of this application were compared with assembled consensus sequences or templates found in the LIFEBSEQ GOLD database. 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.

WO 02/16595

PCT/US01/26317

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 analyze 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 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 $\leq 1 \times 10^{-6}$. The templates were also subjected to frameshift FASTx against GENPEPT, and homolog match was defined as having an E-value of $\leq 1 \times 10^{-6}$. 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/>). Using BLAST2 analysis against the Incyte LifeGold human template database, Clone 3481942CB1 was found to be the 5' end of the 279978.9 template. Further analysis showed that the 3' end was Clone 4331016CT1 within a related template. These two templates were spliced together to form SEQ ID NO:2. 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.

VI 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 KSRCC 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

WO 02/16595

PCT/US01/26317

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.

VII Hybridization Technologies and Analyses

Immobilization of cDNAs on a Substrate

5 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
10 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 37C for 16 hr. The membrane is removed from the agar and consecutively placed colony side up in 10% SDS, 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
15 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 SEPACRYL-400
20 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%
25 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 110C 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
30 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; Tropic, Bedford MA) for 30 min at 60C; 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
35 screening cDNAs, mRNAs, or genomic DNA in membrane-based hybridizations. Probes are

WO 02/16595

PCT/US01/26317

prepared by diluting the cDNAs to a concentration of 40-50 ng in 45 μ l 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 μ l 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 μ l 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 μ l TE buffer and adding 5 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 reaction mixture at ratios of 1:100,000, 1:10,000, 1:1,000, 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 37°C for two hr. The reaction mixture is then incubated for 20 min at 85°C, and probes are purified using two successive CHROMA SPIN+TE 30 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 65°C 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 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 Na₂HPO₄, 5 mM EDTA, pH 7) at 55°C for two hr. The probe, diluted in 15 ml fresh hybridization solution, is then added to the membrane. The membrane is hybridized with the probe at 55°C for 16 hr. Following hybridization, the membrane is washed for 15 min at 25°C in 1mM Tris (pH 8.0), 1% Sarkosyl, and four times for 15 min each at 25°C in 1mM Tris (pH 8.0). To detect hybridization complexes, XOMAT-AR film (Eastman Kodak, Rochester NY) is exposed to the membrane overnight at -70°C, developed, and examined visually.

WO 02/16595

PCT/US01/26317

Polymer Coated Slide-based Hybridization

Probe is heated to 65°C for five min, centrifuged five min at 9400 rpm in a 5415C microcentrifuge (Eppendorf Scientific, Westbury NY), and then 18 µl 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 µl of 5xSSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hr at 60°C. The arrays are washed for 10 min at 45°C in 1xSSC, 0.1% SDS, and three times for 10 min each at 45°C 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 the sample. In the differential hybridization format, differential expression of a set of genes in two 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

WO 02/16595

PCT/US01/26317

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).

VIII 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 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 which show product scores between 8 and 40.

Electronic northern analysis was performed at a product score of 70 as shown in Figure 3. 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 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.

IX 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. These molecules are selected using OLIGO software (Molecular Biology Insights). Detection is described in Example VII. 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 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

WO 02/16595

PCT/US01/26317

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 mammalian protein.

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 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). These 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 the mammalian protein.

15 X Expression of KSRCC

Expression and purification of the mammalian 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 KSRCC in CHO cells. The vector contains the selectable hsd 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 mammalian cDNA 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

XI Production of Antibodies

KSRCC is purified using polyacrylamide gel electrophoresis and used to immunize mice or rabbits. Antibodies are produced using the protocols below. Alternatively, the amino acid sequence of KSRCC 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 coupled to KLH (Sigma-Aldrich) by reaction with N-maleimidobenzoyl-N-

WO 02/16595

PCT/US01/26317

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 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.

XII Purification of Naturally Occurring Protein Using Specific Antibodies

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, the protein is eluted from the column using a buffer of pH 2-3 or a high concentration of urea or thiocyanate ion to disrupt antibody/protein binding, and the protein is collected.

XIII 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.

XIV 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 mammalian 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

WO 02/16595

PCT/US01/26317

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
 5 examined for growth of blue colonies. Interaction between expressed protein and cDNA fusion proteins activates expression of a LBU2 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
 10 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
 15 colonies are isolated and propagated. The pB42AD-cDNA plasmid, which contains a cDNA encoding a protein that physically interacts with the mammalian protein, is isolated from the yeast cells and characterized.

XV KSRCC Assay

AMP-binding activity of KSRCC is determined in a ligand-binding assay using candidate
 20 ligand molecules in the presence of 125 I-labeled KSRCC. KSRCC is labeled with 125 I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate AMP molecules, previously arrayed in the wells of a multi-well plate, are incubated with the labeled KSRCC, washed, and any wells with labeled KSRCC complex are assayed. Data obtained using different concentrations of KSRCC are used to calculate values for the number, affinity, and
 25 association of KSRCC with the candidate molecules.

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 be apparent to those skilled in the art without departing from the scope and spirit of the
 30 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 within the scope of the following claims.

35

WO 02/16595

PCT/US01/26317

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 of the cDNA.
- 5 2. An isolated cDNA comprising a nucleic acid sequence selected from:
 - a) SEQ ID NO:2 or the complement thereof;
 - b) a fragment of SEQ ID NO:2, or the complement thereof; and
 - c) a variant of SEQ ID NO:2 having at least 83% sequence identity to SEQ ID NO:1, or the complement thereof.
- 10 3. A composition comprising the cDNA of claim 1 and a labeling moiety.
4. A vector comprising the cDNA of claim 1.
- 15 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.
- 20 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 under conditions to form at least one hybridization complex; and
 - b) detecting hybridization complex formation, wherein complex formation indicates expression
- 25 of the cDNA in the sample.
8. The method of claim 7 further comprising amplifying the nucleic acids of the sample prior to hybridization.
- 30 9. The method of claim 7 wherein the composition is attached to a substrate.
10. The method of claim 7 wherein complex formation is compared with at least one standard to determine differential expression.
- 35 11. A method of using a cDNA to screen a plurality of molecules or compounds, the method

WO 02/16595

PCT/US01/26317

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 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:

- a) an amino acid sequence of SEQ ID NO:1; and
- b) an antigenic epitope 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 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, 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 an animal with the protein of claim 15 under conditions to elicit an antibody response;
- 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.

WO 02/16595

PCT/US01/26317

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:
- 5 a) combining the antibody of claim 18 with a sample, thereby forming antibody:protein complexes; and
- b) comparing complex formation with a standard, wherein the comparison indicates expression of the protein in the sample.
- 10 20. The method of claim 19 wherein expression is diagnostic of a renal cell carcinoma.
21. A pharmaceutical composition comprising the antibody of claim 18 and a pharmaceutical carrier.
- 15 22. A method for treating a renal cell carcinoma comprising administering to a person in need of such treatment an effective amount of the composition of claim 21.

WO 02/16595

PCT/US01/26317

1/11

```

10 19 28 37 46 55
5' GTGC TCT CCA AGG CTG TAG GAG TTC TGG AGC TGC TGG CTG GAG AGG AGG GTG
---
64 73 82 91 100 109
GAC GAA GGT CTC TCT AGA AAG ACA ACC TGA CAG GAC TTG GCA GGC CTG AAT ATG
---
118 127 136 145 154 163
CAT TGG CTG CGA AAA GTT CAG GGA CTT TGC ACC CTG TGG GGT ACT CAG ATG TCC
---
172 181 190 199 208 217
AGC CGC ACT CTC TAC ATT AAT AGT AGG CAA CTG GTG TCC CTG CAG TGG GGC CAC
---
226 235 244 253 262 271
CAG GAA GTG CCG GCC AAG TTT AAC TTT GCT AGT GAT GTG TGG GAT CAC TGG GCT
---
280 289 298 307 316 325
CAC ATG GAG AAG GCT GGC AAG CGA CTC CCA AGC CCA GCC CTG TGG TGG GTG AAT
---
334 343 352 361 370 379
GGG AAG GGC AAG GAA TTA ATG TGG AAT TTC AGA GAA CTG AGT GAA AAC AGC CAG
---
G K G K E L M W N P R E L S E N S Q

```

Figure 1A

WO 02/16595

PCT/US01/26317

2/11

```

388 CAG GCA GCC AAC GTC CTC TCG GGA GCC TGT GGC CTG CAG CGT GGG GAT COT GTG 433
---
Q A A N V L S G A C G L Q R G D R V
---
442 GCA GTG GTG CTC CCC GGA GTG CCT GAG TGG TGG CTG GTC ATC CTG GGC TGC ATT 487
---
A V V L P R V P E W L V I L G C I
---
496 CGA GCA GGT CTC ATC TTT ATG CCT GGA ACC ATC CAG ATG AAA TCC ACT GAC ATA 541
---
R A G L I F M P G T I Q M K S T D I
---
550 CTG TAT AGG TTG CAG ATG TCT AAG GCC AAG GCT ATT GTT GGT GGG GAT GAA GTC 595
---
L Y R L Q M S K A K A I V A G D E V
---
604 ATC GAA GAA GTG GAC ACA GTG GCA TCT GAA TGT CCT TCT CTG AGA ATT AAG CTA 649
---
I Q E V D T V A S E C P S L R I K L
---
658 CTG GTG TTT GAG AAA AGC TGC GAT GGG TGG CTG AAC TTC AAG AAA CTA CTA AAT 703
---
L V S E K S C D G W L N F K K L L N
---
712 GAG GCA TCC ACC ACT CAT CAC TGT GTG GAG ACT GGA AGC CAG GAA CCA TCT CCC 757
---
E A S T T H H C V E T G S Q E A S A

```

Figure 1B

WO 02/16595

PCT/US01/26317

3/11

```

766 775 784 793 802 811
ATC TAC ACT AGT GGG ACC AGT GGT CTT CCC AAG ATG GCA GAA CAT TCC TAC
I Y F T S G T S G L P K M A E H S Y
---
820 829 838 847 856 865
TCG ACC CTC GGC CTC AAG GGC AAG ATG GAT OCT GGT TGG ACA GGC CTG CAA GGC
S S L G L K A K M D A G W T G L Q A
---
874 883 892 901 910 919
TCT GAT ATA ATG TGG ACC ATA TCA GAC ACA GGT TGG ATA CTG AAC ATC TTG GGC
S D I N W T I S D N G W I L N I L G
---
928 937 946 955 964 973
TCA CTT TTG GAA TCT TGG ACA TTA GGA GCA TGC ACA TTT GTT CAT CTC TTG CCA
S L L E S W T L G A C T F V H L P
---
982 991 1000 1009 1018 1027
AAG TTT GAC CCA CTG GTT ATT CTA AAG ACA CTC TCC AGT TAT CCA ATC AAG AGT
K F D F L V I L K T L S S Y P I K S
---
1036 1045 1054 1063 1072 1081
ATG ATG GGT GCC CCT ATT GTT TAC CCG ATG TTG CTA CAG CAG GAT CTT TCC AGT
N M G A P I V Y R M L L Q Q D L S S
---
1090 1099 1108 1117 1126 1135
TAC AAG TTC CCC CAT CTA CAG AAC TGC CTC GCT GGA GGG GAG TCC CTT CTT CCA
Y K F P H L Q N C L A G G E S L L P

```

Figure 1C

WO 02/16595

PCT/US01/26317

4/11

```

1144 GAA ACT CTG GAG AAC TGG AGG GCC CAG ACA GGA CTG GAC ATC CGA GAA TTC TAT 1171 1180 1189
-----
E T L E N W R A Q T G L D I R E F Y
-----
1198 GGC CAG ACA GAA ACG GGA TTA ACT TGC ATG GTT TCC AAG ACA ATG AAA ATC AAA 1225 1234 1243
-----
G Q T E T G L T C M V S K T M K I K
-----
1252 CCA GGA TAC ARG GGA ACG GGT GCT TCC TGT TAT GAT GTA CAG GTT ATA GAT GAT 1279 1288 1297
-----
P G Y M G T A A S C Y D V Q V I D D
-----
1306 AAG GGC AAC GTC CTG CCC GGC ACA GAA GGA GAC ATT GGC ATC AGG GTC AAA 1333 1342 1351
-----
K G N V L P P G T E G D I G I R V K
-----
1360 CCC ATC AAG CCT ATA GGC ATC TTC TCT GGC TAT GTG GAA AAT CCC CAC AAG ACA 1387 1396 1405
-----
P I R P I G I F S G Y V E N P D K T
-----
1414 GCA CCC AAC ATT CGA GGA GAC TTT TGG CTC CTT GGA GAC CGG GGA ATC AAA GAT 1441 1450 1459
-----
A A N I R G D F W L L G D R G I K D
-----
1468 GAA GAT GGG TAT TTC CAG TTT ATG GGA CGG GCA GAT GAT ATC AAT AAC TTC AGC 1486 1495 1504 1513
-----
E D G Y F Q F M G R A D D I I N S S

```

Figure 1D

WO 02/16595

PCT/US01/26317

5/11

```

1522      1531      1540      1549      1558      1567
GGG TAC CGG ATT GGA CCC TCG GAG GTA GAG AAT GCA CTG ATG AAG CAC CCT GCT
---
G   Y   R   I   G   P   S   E   V   E   N   A   L   M   K   H   P   A
---
1576      1585      1594      1603      1612      1621
GTG GTT GAG ACG GCT GTG ATC ACC AGC CCA GAC CCC GTC CGA GGA GAG GTG GTG
---
V   V   E   T   A   V   I   S   S   P   D   P   V   R   G   E   V   V
---
1630      1639      1648      1657      1666      1675
AAG GCA TTT GTG ATA CTG GCC TCG CAG TTC CTA TCC CAT GAC CCA GAA CAG CTC
---
K   A   P   V   I   L   A   S   Q   F   L   S   H   D   P   E   Q   L
---
1684      1693      1702      1711      1720      1729
ACC AAG GAG CTG CAG CAG CAT GTG AAG TCA GTG ACA GCC CCA TAC AAG TAC CCA
---
T   K   E   L   Q   Q   H   V   K   S   V   T   A   P   Y   K   Y   P
---
1738      1747      1756      1765      1774      1783
AGA AAG ATA GAG TTT GTG TTG AAC CTG CCC AAG ACT GTC ACA GGG AAG ATT CAA
---
R   K   I   E   F   V   L   N   L   P   K   T   V   T   G   K   I   Q
---
1792      1801      1810      1819      1828      1837
CGA ACC AAA CTT CGA GAC AAG GAG TGG AAG ATG TCC GGA AAA GCC CGT GCG CAG
---
R   T   K   L   R   D   K   E   W   K   M   S   G   K   A   R   A   Q
---
1846      1855      1864      1873      1882      1891
TGA GGC GTC TAC GAG ACA TTC ATT TGG ATT CCC CTC TTC TTT CTC TTT CTC TTC
---
+

```

Figure 1E

WO 02/16595

PCT/US01/26317

6/11

1900	1909	1918	1927	1936	1945
CCG TTG GGC CCT TGG CCT TAC TAT GAT GAT ATG AGA TTC TTT ATG AAA GAA CAT					
1954	1963	1972	1981	1990	1999
GAA TGT AAG TTT TGT CTT GCC CNG GTT ATT AGC ACA AAA CAT TAC TAT GTT AGA					
2008	2017	2026	2035	2044	2053
TAT TGA AAT AAG AAA GGG AAG GAA TGA GAG AGA GAG AAA AGG AGA GGG TAA CAG					

A 3'

Figure 1F

WO 02/16595

PCT/US01/26317

7/11

```

1  MHLLW  K I PRUCTEM  ETEM  HRS  EFMNIAK  LMP  IQWGHORV  93127193
1  M  -----  93219339
1  MHLLRKVOGLC  TLNGTQ  M  SRT  V  LNSROLV  SLOWGHOEV  KSRCC
41  PAKFN  FASDV  IDMA  SL  E  AGKRS  P  P  A  H  MNNGS  GEF  L  S  93127193
2  -----  93219339
41  PAKFN  FASDV  IDMA  SL  E  AGKRS  P  P  A  H  MNNGS  GEF  L  S  KSRCC
61  K  F  R  E  L  S  E  I  S  G  E  F  N  V  L  T  A  C  G  L  O  R  D  R  V  A  V  L  F  R  Y  P  E  W  93127193
2  -----  93219339
61  K  F  R  E  L  S  E  I  S  G  E  F  N  V  L  T  A  C  G  L  O  R  D  R  V  A  V  L  F  R  Y  P  E  W  KSRCC
81  K  F  R  E  L  S  E  I  S  G  E  F  N  V  L  T  A  C  G  L  O  R  D  R  V  A  V  L  F  R  Y  P  E  W  93127193
2  -----  93219339
81  K  F  R  E  L  S  E  I  S  G  E  F  N  V  L  T  A  C  G  L  O  R  D  R  V  A  V  L  F  R  Y  P  E  W  KSRCC
121  L  V  T  L  G  C  M  F  S  G  L  V  F  P  G  T  T  Q  K  S  T  D  I  Y  R  L  O  S  S  K  A  R  A  I  Y  A  G  93127193
2  -----  93219339
121  L  V  T  L  G  C  M  F  S  G  L  V  F  P  G  T  T  Q  K  S  T  D  I  Y  R  L  O  S  S  K  A  R  A  I  Y  A  G  KSRCC
161  D  E  V  I  O  E  V  D  T  V  A  S  E  C  P  S  L  R  I  K  L  V  S  E  K  S  C  D  G  W  L  N  E  K  K  L  N  E  931271
2  -----  932193
161  D  E  V  I  O  E  V  D  T  V  A  S  E  C  P  S  L  R  I  K  L  V  S  E  K  S  C  D  G  W  L  N  E  K  K  L  N  E  KSRCC
201  A  S  P  I  H  B  D  V  E  I  V  S  O  E  S  A  A  I  Y  F  T  S  G  T  S  G  P  P  F  M  A  E  H  S  M  C  S  I  G  L  931271
2  -----  932193
201  A  S  T  T  H  C  V  E  T  G  S  O  E  A  S  A  I  Y  F  T  S  G  T  S  G  L  P  K  M  A  E  H  S  Y  S  S  L  G  L  KSRCC
241  K  A  K  M  O  A  G  W  T  G  L  G  W  S  D  T  K  M  T  I  S  D  T  G  W  L  N  I  L  G  S  P  L  E  P  V  L  G  931271
2  -----  932193
241  K  A  K  M  O  A  G  W  T  G  L  G  W  S  D  T  K  M  T  I  S  D  T  G  W  L  N  I  L  G  S  P  L  E  P  V  L  G  KSRCC
281  T  Q  I  F  V  H  L  L  K  F  P  P  O  T  V  L  V  I  S  S  Y  P  I  N  T  L  G  A  P  L  I  Y  R  M  L  L  O  93127193
2  -----  93219339
281  A  C  T  F  V  H  L  L  K  F  P  P  O  T  V  L  V  I  S  S  Y  P  I  N  T  L  G  A  P  L  I  Y  R  M  L  L  O  KSRCC
321  Q  D  L  S  S  Y  K  F  P  P  L  H  S  C  E  S  G  E  T  L  E  F  T  L  E  S  W  F  A  K  T  E  L  F  A  F  A  93127193
2  -----  93219339
321  Q  D  L  S  S  Y  K  F  P  P  L  H  S  C  E  S  G  E  T  L  E  F  T  L  E  S  W  F  A  K  T  E  L  F  A  F  A  KSRCC

```

Figure 2A

PCIT/US01/26317

Figure 2B

560	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931271193
561										NSMCC
562	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931293339
563										NSMCC
564	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931271193
565										NSMCC
566	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931293339
567										NSMCC
568	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931271193
569										NSMCC
570	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931293339
571										NSMCC
572	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931271193
573										NSMCC
574	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931293339
575										NSMCC
576	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931271193
577										NSMCC
578	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931293339
579										NSMCC
580	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931271193
581										NSMCC
582	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931293339
583										NSMCC
584	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931271193
585										NSMCC
586	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931293339
587										NSMCC
588	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931271193
589										NSMCC
590	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931293339
591										NSMCC
592	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931271193
593										NSMCC
594	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931293339
595										NSMCC
596	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931271193
597										NSMCC
598	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931293339
599										NSMCC
600	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931271193
601										NSMCC
602	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931293339
603										NSMCC
604	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	931271193
605										NSMCC
606	VGOTETG	ICN	SR	KKV	GYL	IA	IV	YDVQVID	EGM	9

WO 02/16595

PCT/US01/26317

9/11

Found in:

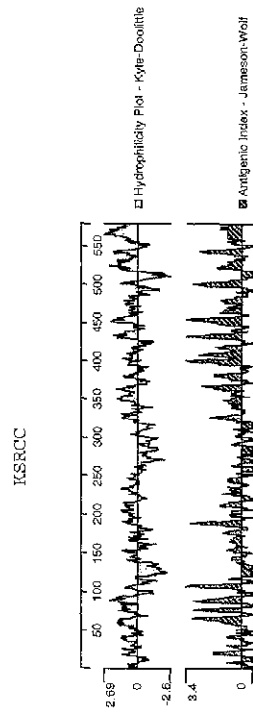
Library	Clone	Library Description	Abs	Pct	Pct
ID	Count		Abundance		Spec
KIDN00132	5627	Kidney, 42M	4	0.0711	8.62
KIDN00125	3810	Kidney, mw/benign cyst, nephrolithiasis, 42F	2	0.0525	6.16
KIDN00105	2352	Kidney tumor, renal cell, 3, CGAP	1	0.0425	4.91
KIDN00103	3141	Kidney, mw/renal cell CA, 8.53F, pool, NORM	1	0.0318	3.72
KIDN00126	3299	Kidney, medulla/cortex, mw/renal cell CA, 53F, n/KIDN00116	1	0.0303	3.55
KIDN00119	6963	Kidney, mw/renal cell CA, 65M n/KIDN00115	2	0.0287	3.46
KIDN00101	6149	Kidney, cortex, mw/renal cell CA, 65M	1	0.0163	1.94
KIDN00102	1869	Kidney, cortex, mw/renal cell CA, 65M	1	0.0535	24.53
KIDN00131	3515	Kidney	1	0.0284	14.73
KIDN00115	2954	Kidney tumor, renal cell CA, 65M n/KIDN00119	1	0.0253	13.31

Figure 3

WO 02/16595

PCT/US01/26317

10/11



g3127193

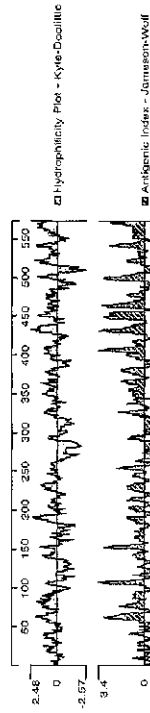


Figure 4A

WO 02/16595

PCT/US01/26317

11/11

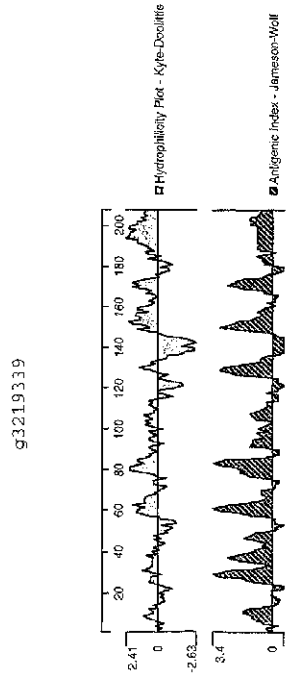


Figure 4B

WO 02/16595

PCT/US01/26317

<110> INCYTE GENOMICS, INC.
WALKER, Michael G.
KRASNOW, Randi E.

<120> KIDNEY-SPECIFIC PROTEIN

<130> PC-0020 OCT

<140> To Be Assigned
<141> Herewith

<150> 09/645,961

<151> 2000-08-24

<160> 16

<170> PERL Program

<210> 1

<211> 577

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> incyte ID No: 3481942

<400> 1

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

WO 02/16595

PCT/US01/26317

Gly Ser Leu Leu 260 265 270
 Glu Ser Trp Thr Leu Gly Ala Cys Thr Phe Val
 His Leu Leu Pro Lys Phe Asp Pro Leu 280 285
 Val Ile Leu Lys Thr Leu
 Ser Ser Tyr Pro Ile Lys Ser Met Met Gly Ala Pro Ile Val Tyr 290 300
 305 310
 Arg Met Leu Leu Gln Gln Asp Leu Ser Ser Tyr Lys Phe Pro His 315 320
 325 330
 Leu Gln Asn Cys Leu Ala Gly Gly Glu Ser Leu Leu Pro Glu Thr 335 340
 345 350
 Leu Glu Asn Trp Arg Ala Gln Thr Gly Leu Asp Ile Arg Glu Phe 355 360
 365 370
 Tyr Gly Gln Thr Glu Thr Gly Leu Thr Cys Met Val Ser Lys Thr 375 380
 385 390
 Met Lys Ile Lys Pro Gly Tyr Met Gly Thr Ala Ala Ser Cys Tyr 395 400
 405 410
 Asp Val Gln Val Ile Asp Asp Lys Gly Asn Val Leu Pro Pro Gly 415 420
 425 430
 Thr Glu Gly Asp Ile Gly Ile Arg Val Lys Pro Ile Arg Pro Ile 435 440
 445 450
 Gly Ile Phe Ser Gly Tyr Val Glu Asn Pro Asp Lys Thr Ala Ala 455 460
 465 470
 Asn Ile Arg Gly Asp Phe Trp Leu Leu Gly Asp Arg Gly Ile Lys 475 480
 485 490
 Asp Glu Asp Gly Tyr Phe Gln Phe Met Gly Arg Ala Asp Asp Ile 495 500
 505 510
 Ile Asn Ser Ser Gly Tyr Arg Ile Gly Pro Ser Glu Val Glu Asn 515 520
 525 530
 Ala Leu Met Lys His Pro Ala Val Val Glu Thr Ala Val Ile Ser 535 540
 545 550
 Ser Pro Asp Pro Val Arg Gly Glu Val Val Lys Ala Phe Val Ile 555 560
 565 570
 Leu Ala Ser Gln Phe Leu Ser His Asp Pro Glu Gln Leu Thr Lys 575
 580
 Glu Leu Gln Gln His Val Lys Ser Val Thr Ala Pro Tyr Lys Tyr
 Pro Arg Lys Ile Glu Phe Val Leu Asn Leu Pro Lys Thr Val Thr
 Gly Lys Ile Gln Arg Thr Lys Leu Arg Asp Lys Glu Trp Lys Met
 Ser Gly Lys Ala Arg Ala Gln
 595

<210> 3

<211> 2054

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3481942

<400> 2

gtgctctctt ccaaggctgt agggattctg gagctgctgg ctggagagga gggtagcga 60
 agctctctct agaaagacat cctgagagga cttggcaggg ctgcatatgc attgctgcg 120
 aaaaattccag gaactttgca cctgtgggg tactccagatg tccagccgca ctctctaat 180
 taatagtaga caactggatg cctgtgagtg gggccaccag gaagtgccgg caagtttaa 240
 ctgtgctagt gatgtgttg atcactgggc tgaatggag aaggttgcca agcgactcc 300
 aagccagacc ctgtgtgggg tgaatggcaa ggggaaggaa ttaattgga atttcagaga 360
 actgagtgaa acagccatgg aggcagccaa cttctctctg ggagcctgk gcttcagcgg 420
 tggggatcgt gtggcagtg tgctgcccgc agtgcctgag tggtagctgg tgatccctgg 480
 ctgcattcga gcaggtctca tctttatgcc tggaccatcc cagatgaat caactgcat 540
 actgtatagg ttgcagatgt ctasggccaa ggctattgtt gctggggatg agtcatcca 600
 agaagtggac acagtggcat ctgaatgtcc ttctatgaga attagctac tgggtgctga 660

WO 02/16595

PCT/US01/26317

```

gaaaagctgc gatgggtggc tgaacttcaa gaaactacta aatgaggcat ccaccactca 720
tcactgtgtg gagactggaa gccaggaaac atctgccatc tacttcacta gtgggaccag 780
tggctctccc aagatggcag aacattccta ctcgagccctg ggcttcacgg ccagatggga 840
tgctgtgttg aagggctgc agcctcttga tataatgttg aacatacag acacaggltg 900
gtactctgac atctggggtc aacttttggg atcttggaaa ttaggagcat gcaactttgt 960
tcactctctg ccaagtttg aacctctgtt tattctaaa acactctcca gttatccaat 1020
caagagtatg atgggtggcc ctattgttta cgggatgttg ctacagcagg atcttccag 1080
ttacagtttc cccattctac agaactgctc cgtggagggg gactcccttc ttccagaaac 1140
tcaggagAAC tggaggggcc agacaggact ggacatccga gaattctatg gccagacaga 1200
aacgggatta acttgcatgg ttccaagac aatgaaatc aaacccagat acatgggAAC 1260
ggctgtcttc tgttatgatg tacaggttat agatgataag ggcacagtc ttcccccg 1320
ccagaggaag gacattggca tcagggtcaa accatcagg cctataggca tctctctcgg 1380
ctatgtgaa aatccccaca agacagcagc caacattcga ggagactttt ggtctctcgg 1440
agacgggga stcaagatg agatgggta ttccagttt atgggacggg cagatgatat 1500
cattactcc agcgggAAC ggaattggac ctgggaggla gagaatgcac tgaLgaagca 1560
ccctgtgtgt gttgagaggg ctgtgtcag cagccagac ccgtccgag gagaggttgt 1620
gagggaatt gtgatactgg cctcgagtt cctalcccat gacccagAAC agctcaccaa 1680
ggagctgag cagcatgtga agtcagtgac agcccatc acgtaccca gaaagtcga 1740
gtttgtcttg accctgcca agactgtac agggaaatt caacgaacca aacttcgaga 1800
caaggagtg agatgtccg gaaaagcccg tggcagtgga ggcgtctagg agaatctcat 1860
ttggatcccc ctctctcttc tcttctcttc ccttttggg ccttgccctt actatgatga 1920
tatgagatcc ttatggaag aacatgaatg taagtlttgt ctbgccctgg ttattagcac 1980
aacaattuc tatgttgtat attgaaataa gaaagggag gatgagaga gagtgaagaa 2040
gaggggtaa caga

```

```

<210> 3
<211> 521
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 3481942CB1

```

```

<400> 3
totggagctg ctggctggag aggggggtgg acgaagctct ctctagaaag acatccctgag 60
aggactttgc agccctgaac atgcattggc tgcgaaaagt tcagggaact tgcacctgt 120
ggggtautca gatgtccagc cgcactctct acattaatag taggcaactg gtgtccctgc 180
agtggggcca ccagggaagt ccggccaagt ttaactttgc tagtgatgtg ttggatcaat 240
gggtgacat ggaagaagct ggcaagcagc tcccaagccc agccctgttg tgggtgaatg 300
ggaaggggaa ggaattaatg tggatttcca gagaactgag tgaacacagc cagcaggcag 360
ccaaagctct ctggggagcc tctggcctgc agcgtgggga tctgtggcca gtggtctcgc 420
ccggagtcac tgaatgtgtg ctgggtgac ttgggtgcat tggagcaggg ctcatcttta 480
tgcttggaac catcagatga gatccactga catactgtat a 521

```

```

<210> 4
<211> 251
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 5519427H1

```

```

<220>
<221> unsure
<222> 25, 55
<223> a, c, g, or other

```

```

<400> 4
gtgctctctt ccaagctgtt aggaagtctt ggaagtctgt gctggagagg agggmggagc 60
aagctctctc cagaaagaca tctggagagg acttggcagg cctgaacatg cattggctgc 120
gaaaagtcca gggactttgc accctgtggg gtaactcagat gtccagccgc actctctaca 180
ttaatagtag gcaactggtg tccctgcagt ggggccaaca ggaattccgg ccaagttaa 240

```

WO 02/16595 PCT/US01/26317

ctttgctagt g 251

<210> 5
 <211> 594
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 539098478

<400> 5
 ggctgttaga gttctggagc tgcaggctgg agaggagggt ggaatgaagct ctctctagaa 60
 agacatcctg agaggacttg gcaggcctga atatgcattg gctgcgaasa gtccaggac 120
 ttctgaacct gtgggggtact cagatgtcca gcgcactct ctacattaat agtaggcac 180
 tgggtgacct gcagtggggc caccaggagc tgcgggcaaa gtttaacctt gctagtgtgt 240
 tgttgatca ctggcctgac atggagaagg ctggcaagcg actcccaagc ccagccctgt 300
 gttgggtgaa tgggaaggag agggatttaa tctggattt cagggaactg cgtgaaaaa 360
 gccagcagcg agccaactc ctctgggag cctgtggcct gcgcgtggg gctgtgtgg 420
 cagtgtgtct gcccgagtg cctgagtggt ggcctgtgat cctgggtctg attcagcag 480
 gtctatcttt tatgcctgga accatccaga tgaatccac tgacatactg tatagtgctg 540
 agatgtctaa ggccaaggct attgtgtctg gggatgaagt catccagaa gtgg 594

<210> 6
 <211> 444
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 433223079

<400> 6
 gttgatctaa agacctttag aataaccagt gggtcaact ttggcaagag atgaacaaat 60
 gtgcctgctc ctaatgtcca agattccaaa agtgagccca agatgttcag tatccaaact 120
 gtgtctgata tggtcacat tatatcagag gcttgacggc ctgtccaaac agcatcact 180
 ttggccttga ggcacaggct cagtaggaa tgttctgcca tcttggaag accactggtc 240
 ccactagtga agtagatggc agatgttcc tggcttccag tctccacaa gtgatgagt 300
 ctgatgctt catttagtag ttcttgagc ttcagccccc catcgagct ttctcagac 360
 accagtgcct taattctcag agaaggacat tcagatgcca ctgtgtccac ttcttggtat 420
 acttcatccc cagcaacaat agcc 444

<210> 7
 <211> 535
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 760835K1

<400> 7
 atgtggacca tatcagcac aggttgata ctgaacatct tgggctcact tttggaaact 60
 tggacattag gagcatgccc atttgttcat ctcttgccaa agtttgacc actggttatt 120
 ctcaagacac tccacagtta tccaatcaag agtatgatag atgcccctat tatctacag 180
 atgttgatag agcaggatct ttccagttac aggttcccc atttaccaga ctgctcgtct 240
 ggaaggaggt ccttcttcc agaaactctg gagaactgga gggcccaaac aggaactgac 300
 atcaggaatc tctatggcca gacagaacgc ggattaaactt gatgggttc caagcaaatg 360
 aaaaatcaac caggatcatc gggaaaggct gcttctgtt algatgtaca ggttatagt 420
 gataaggcca acgtcctgccc ccccgccaca gaaggagaca ttggcctcag ggttaaaccc 480
 atcaggccca taggcacttt ctctggctat gtggtatcat ccccatgac tatgc 535

<210> 8
 <211> 487

WO 02/16595

PCT/US01/26317

```

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 762540R1

<400> 8
agtccccca tctacagagc tgctctgctg gaggggagtc ccttcttcca gaaactctgg 60
agaactggag ggcccagaca ggaactggaca tccgagaatt ctatggccag acagaaacgg 120
gattaaacttg catggtttcc aagacaatga aaatcaaac accgtacatg ggaacggctg 180
cttctctgtra tgatgtacag gttatagatg ataagggcac cgtctctgcc ccggccacag 240
aaggagacat tggcatcagg gtaaaaccca tcaggccctat aggcactcttc cctggctatg 300
tggaaatcc cgaacagaca gcagccaaca ttctggagga ctttttggctc cttggagacc 360
ggggaaatcaa agatgaagat gggttatttc agtttatggg acgggcagat gatattccta 420
actccacggg gtaaggattg gacctctgga gatagagaat gcactgatga agcaccctgc 480
tgtggatt

<210> 9
<211> 488
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2099420R6

<400> 9
caacattcca ggaagactttt ggtctcttgg aagccgggga atcaaatgat aagatgggta 60
ttccagattt atggagcggg cagatgatat catlaattcc agcgggtacc agattggacc 120
ctggagggta gagaatgac tgatgaaga cctgtctgtg gttgagacgg ctgtgacag 180
cagggcagac cccgtccgag gaaagctggt gaagcatttt gtgatacttg cctccagatt 240
cctatcccat gcccagaaac agctaccca gaaactctg cagcatatga agtcaatgac 300
agcccatcac aagtaccaca gaagataga gtttgtcttg aacctgccca agactgtcac 360
agggaaatatt caacgaacca aaactccaga caaggagtgg aagatgtccg gaaagacccg 420
tcgcagtgga ggcgtctagg agacattcat ttggattccc ctcttctttc tcttttcttt 480
ccctcttg

<210> 10
<211> 528
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7695905J1

<400> 10
ggggtgggta aggcatttgt gatactggcc tcgcagttcc tatccatga cccagaacag 60
ctcaccagag agctgcagca gcatgtgaag tcagtgcacg ccccatacaa gtacccaaga 120
aagatagagt ttgtcttgaa cctgtccaa agtgcacag ggaacattca acgaaccaa 180
cttcagagaca aggagtggaa gctgtccgga aaagcccggt gcagtgagg cgtctaggag 240
acattcattt ggattccctt ctctctcttc ttcttttttc ctttgggcc tttgacctac 300
tatgatgata tgagattctt tatgaagaa catgaatgta agttttgtct tgccttggtt 360
attagcacaa accatttcta tcttagatat tgaataaga aagggaagga atgagagaga 420
gtgaagagga gagggtataa gaaaagagg aaagaaagt aagtcaggga aatattaaaa 480
actgcaaggg aaggaattg aaaaagaat aaagtatgga ggggaaggc 528

<210> 11
<211> 310
<212> DNA
<213> Rattus norvegicus

<220>

```

WO 02/16595

PCT/US01/26317

```

<221> misc_feature
<223> Incyte ID No: 701648693K1

<400> 11
gtttcaggga gctgagggaa cttagcgccc gtgctgcaaa cgtctttgag cagacttggt 60
gctctgagaa tggagatggc ctggccctga tcttgccctg agtgcgggaa tcttgggtga 120
tgcagctggg ctgcattcga acaggagctca tcttcattcc tgggautgac caaatgaaag 180
ccaaaggatat tctctacaga atacaatatgt ctcaagccaa agccatttgt accacagaca 240
gccttcttcc agaggtggaa tctgtggctt ctgagtgctc tggcttgaaa accaagatag 300
tgggtgtctga                                     310

<210> 12
<211> 358
<212> DNA
<213> Rattus norvegicus

<220>
<221> misc_feature
<223> Incyte ID No: 212451_Rn.1

<400> 12
atccccoatt catcagtggt tggagactgt aagccaagaa tcagctgcca tctatttcac 60
tagcggggacc agtggccctc ccaagatggc agagcactcc cactgcagcc tgggcctcaa 120
ggccaaagatg gatgctggct ggacaggact gggacottct gacacatgt gggaccatctc 180
agacacaggc tggatatcaa acatttttagg gtcatctctg gacotttggg tatttgggaac 240
gtgcataatt gtccctcttt tgcacaagtt tgatccacaa actggtctaa aggtgcttctc 300
cagctacccc atcaataccc tgcctgggtgc cccctcoatt taccggatgt tctacaaa 358

<210> 13
<211> 1428
<212> DNA
<213> Rattus norvegicus

<220>
<221> misc_feature
<223> Incyte ID No: 203264_Rn.1

<400> 13
attcgaaacg aggaacagga tctttccagt tacaagtccc cacatctgca tagctgtctc 60
agtggaggag agaccctcct cccggugact ctggagagtt ggcagcgcaa gacaggactg 120
ggaaatccag aaatctatgg ccagacagaa acgggaatta cctgcagagt ttctaggaca 180
atgaaagtca aaccaggcta cctgggaaca gccattgtcc ctatgatgt ccagggtcata 240
gatggcgagg gcaatgtcct gccccctggc aaggaaggag acatggctctc caggggtgaag 300
cccatcaggc ctataggcat gtctctcggg tatgtggaca atcccaagaa gacacaggctc 360
aaatctcag gaaacttttg gcttctggga gacgggggaa ttaaggtac agaaaggtatc 420
ttcaacttca tgggaaggaa agaagatctc attaatcca gtgggtaccg gattgggacc 480
ttcaggggtg ggaatgnaa tgatggaaca tctgcctgt gtgaaacag ctgtgatcag 540
cagccacagc ctatcagaa gagaggtggt gaaggcatit gtggttctag cccctgagtt 600
cctgtcccat gacaaagacu agctacccaa ggttcttcag gaacacgtga agtcaagtac 660
agcacccatc aegtacccca ggaagggtga gtttgtctta gacotgcca agaccatcat 720
caggaaaaat tgagcagatc aaaccttcga gccaaaggaat ggaataacatc aggtatagcc 780
caggcccaag tgagactccc aaggctcttg ctctgtctgt tcccaatcct tttcttatga 840
ctccttagtc ttctctatgc aatctgaaat tattctatgt agagggcatgt atgtgattta 900
ggtctttgct tgggttatgt actactagaa cagatgacac gtttctatgc taaaagatag 960
tcacagcatg agggataaag aggtttgaag tctagaaaag gctaaaacag caaaggggaa 1020
agaaattgcc aattcatgtg gtccggaaat ggttagagag aaggggatat agcagaatga 1080
gggaagaagc agaaagatga gaatggaaatc tggagagaaa tagtataaaa gggattgatg 1140
tagaaggaca ctagtgtcat tgcacagaga ttgtgtcaat ccttagggac atcatgaccc 1200
aaagagaaat ggaataacca cttaacagac atgacccagg gcaactgtat tataacttta 1260
tcattgattcc atatccccc cgtacatgag ccattcttaca ttcaataatg ggtgaactct 1320
gcattggctca tcatctatgc accccaataa tgtatcaaca ttctgttagc tcaatatttt 1380
tgttcttagc ttatttctct gaattaataa taaatcactc tcaaaatc 1428

<210> 14

```

WO 02/16595

PCT/US01/26317

<211> 464
 <212> DNA
 <213> *Rattus norvegicus*

<220>
 <221> misc_feature
 <223> Incyte ID No: 701836346T1

<400> 14
 ttctggcacc gtccgcttct ctctctctct taccctctcc tgaccacatg aattgggcaat 60
 ttctttccct tgcgtgttta gctttttctg acttcaaac tctttttcc ttctttttc 120
 ttcttttagc atgcaacgtg tcattgttct gtgtcaataa ccaagcaag aactaatca 180
 cctacatgcc tctacatgaa taatttcata ttgtatagg aagactaagg agtcataaga 240
 aaaggatttg gaagacaga gcaagagcct tgggagcttc actgggcttg ggttatcct 300
 gatgttttcc attcttggc tcgaagttta gctcgctcaa tttttctgt gatggtcttg 360
 ggccggttca agacaaactc cacttctctg ggtacttgt aggtgtctgt cactgacttc 420
 acgtattcct gaagacactt ggtgagcttg tcttggtcat gggg 464

<210> 15
 <211> 572
 <212> PRT
 <213> *Rattus norvegicus*

<220>
 <221> misc_feature
 <223> Incyte ID No: g3127193

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

WO 02/16595

PCT/US01/26317

Gly Ser Phe Leu Glu Pro Trp Val Leu Gly Thr Cys Ile Phe Val
 275 280 285
 His Leu Leu Pro Lys Phe Asp Pro Gln Thr Val Leu Lys Val Leu
 290 295 300
 Ser Ser Tyr Pro Ile Asn Thr Leu Leu Gly Ala Pro Leu Ile Tyr
 305 310 315
 Arg Met Leu Leu Gln Gln Asp Leu Ser Ser Tyr Lys Phe Pro His
 320 325 330
 Leu His Ser Cys Phe Ser Gly Gly Glu Thr Leu Leu Pro Glu Thr
 335 340 345
 Leu Glu Ser Trp Lys Ala Lys Thr Gly Leu Glu Ile Arg Glu Ile
 350 355 360
 Tyr Gly Gln Thr Glu Thr Gly Ile Thr Cys Arg Val Ser Arg Thr
 365 370 375
 Met Lys Val Lys Pro Gly Tyr Leu Gly Thr Ala Ile Val Pro Tyr
 380 385 390
 Asp Val Gln Val Ile Asp Glu Gln Gly Asn Val Leu Pro Pro Gly
 395 400 405
 Lys Glu Gly Asp Met Ala Leu Arg Val Lys Pro Ile Arg Pro Ile
 410 415 420
 Gly Met Phe Ser Gly Tyr Val Asp Asn Pro Lys Lys Thr Gln Ala
 425 430 435
 Asn Ile Arg Gly Asp Phe Trp Leu Leu Gly Asp Arg Gly Ile Lys
 440 445 450
 Asp Thr Glu Gly Tyr Phe His Phe Met Gly Arg Thr Asp Asp Ile
 455 460 465
 Ile Asn Ser Ser Gly Tyr Arg Ile Gly Pro Ser Glu Val Glu Asn
 470 475 480
 Ala Leu Met Glu His Pro Ala Val Val Glu Thr Ala Val Ile Ser
 485 490 495
 Ser Pro Asp Pro Ile Arg Arg Glu Val Val Lys Ala Phe Val Val
 500 505 510
 Leu Ala Pro Glu Phe Leu Ser His Asp Gln Asp Gln Leu Thr Lys
 515 520 525
 Val Leu Gln Glu His Val Lys Ser Val Thr Ala Pro Tyr Lys Tyr
 530 535 540
 Pro Arg Lys Val Glu Phe Val Leu Asp Leu Pro Lys Thr Ile Thr
 545 550 555
 Gly Lys Ile Glu Arg Ala Lys Leu Arg Ala Lys Glu Trp Lys Thr
 560 565 570
 Ser Gly

<210> 16

<211> 207

<212> PPT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: g3219339

<400> 16

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

WO 02/16595

PCT/US01/26317

```

Arg Ala Asp Asp Ile Ile Asn Ser Ser Gly Tyr Arg Ile Gly Pro
          95          100          105
Ser Glu Val Glu Asn Ala Leu Met Glu His Pro Ala Val Val Glu
          110          115          120
Thr Ala Val Ile Ser Ser Pro Asp Pro Val Arg Gly Glu Val Val
          125          130          135
Lys Ala Phe Val Val Leu Ala Leu Gln Phe Leu Ser His Asp Pro
          140          145          150
Glu Gln Leu Thr Lys Glu Leu Gln Gln His Val Lys Ser Val Thr
          155          160          165
Ala Pro Tyr Lys Tyr Pro Arg Lys Ile Glu Phe Val Leu Asn Leu
          170          175          180
Pro Lys Thr Val Thr Gly Lys Ile Gln Arg Ala Lys Leu Arg Asp
          185          190          195
Lys Glu Trp Lys Met Ser Gly Lys Ala Arg Ala Gln
          200          205

```

【国際公開パンフレット（コレクトバージョン）】

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

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
28 February 2002 (28.02.2002)

PCT

(10) International Publication Number
WO 02/016595 A3

(51) International Patent Classification: C12N 15/00

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

(21) International Application Number: PCT/US01/26317

(22) International Filing Date: 23 August 2001 (23.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/645,981 24 August 2000 (24.08.2000) US

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 5150 Porter Drive, Palo Alto, CA 94304 (US);

(81) Designated States (national): AU, AG, AI, AM, AT, 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, GR, 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, MY, NO, NZ, PL, PT, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GI, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— with international search report

[Continued on next page]

(54) Title: KIDNEY-SPECIFIC PROTEIN

```

5' CTGC TCT CTT CGA AGG CAG CAG GAT TCC TGG AGT TGC TGC CTG CAG AGC AGG CAG
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995

```

(57) Abstract: The invention provides a mammalian cDNA which encodes a kidney specific 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 kidney disorders.

WO 02/016595 A3

WO 02/016595 A3



(88) Date of publication of the international search report: 16 January 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

【国際調査報告】

INTERNATIONAL SEARCH REPORT		Int. Application No. PCT/US 01/26317
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/00		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Magnetic documentation searched (classification system followed by classification symbols) IPC 7 C12N C07K		
Documentation searched other than magnetic documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EMBL, SEQUENCE SEARCH		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HILGERS K.F. ET AL.: "Molecular cloning of K5, a novel rat gene expressed exclusively in the kidney" KIDNEY INTERNATIONAL, vol. 54, 1998, pages 1444-1454, XP002283261 the whole document ---	1-22
☒ Further documents are listed in the continuation of box C. ☒ Patent family members are listed in annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "C" earlier document but published on or after the international filing date "D" document which may know (could be a priority claim) or which is cited to assist in the publication date of another claim or other special reason (see description) "E" document referring to an oral disclosure, use, exhibition or other means "F" document published prior to the international filing date but later than its priority date claimed "G" later document published after the international filing date or priority date and not in conflict with the applicant's disclosure in understanding the principles or theory underlying the invention "X" document of particular relevance, the claimed invention is not disclosed therein or does not be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance, the claimed invention is not disclosed therein or does not be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "Z" document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
26 June 2002		11 07 2002
Name and mailing address of the ISA European Patent Office, P.O. Box 9360 CH-2000 Lausanne Tel: (+41-76) 310 2040, Fax: 31 65 1 60 01 Fax: (+41-76) 310 2040		Authorized officer Macchia, G

IPC Class. (IPC Class) (July 1989)

INTERNATIONAL SEARCH REPORT		Title PCT/US 01/26317
C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] 1 November 1998 (1998-11-01) "Homolog of rat kidney-specific (KS) gene, A-923A4.1, Homo sapiens" Database accession no. 075202 XP002203262 the whole document -8 LOFTUS B.J. ET AL.: "Genome duplications and other features in 12 Mb of DNA sequence from human chromosome 16p and 16q" GENOMICS, vol. 60, 1999, pages 295-308, XP002167778 ---	1-22
X	WO 98 45437 A (GENETICS INSTITUTE INC [US] JACOBS MCCOY LAVALLIE RACIE MERBERG TRECAY) 15 October 1998 (1998-10-15) SEQ ID NO:987 abstract page 56 -page 85 ---	2,13-19, 21
X	DATABASE EMBL [Online] 3 February 1999 (1999-02-03) "Homo sapiens cDNA clone IMAGE:1914672 3' similar to TR:Q13732" Database accession no. A1310615 XP002203264 the whole document ---	2
X	DATABASE EMBL [Online] 4 March 2000 (2000-03-04) "Homo sapiens cDNA clone IMAGE:1492165 5' similar to TR:070490 kidney-specific protein" Database accession no. A1791560 XP002203263 the whole document ---	2
P.X	WO 00 58473 A (CURAGEN CORPORATION (US); SHIMKETS RICHARD A.; LEACH MARTIN) 5 October 2000 (2000-10-05) ORF3009; SEQ ID NO:6017, 6018 abstract page 386 page 5193 -page 5195 claims ---	2,13-19, 21
E	WO 01 61622 A (INCYTE PHARMACEUTICALS INC. (US) WALKER VOLKMUTH KLINGLER AZIMZAI YUE) 23 August 2001 (2001-08-23) SEQ ID NO:3, 10 abstract -----	2,13-22

Form PCT/US 01/01 (Annex to the PCT Rules) July 1992

INTERNATIONAL SEARCH REPORT	
	International application No. PCT/US 01/26317
Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)	
This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:	
1.	☒ Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely: Although claim 22 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2.	☐ Claims Nos.: because they relate to parts of the International Application that do not comply with the prescriptive requirements to such an extent that no meaningful International Search has been carried out, specifically:
3.	☐ Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(c).
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:	
1.	☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2.	☐ As all searchable claims could be searched without effort involving an additional fee, this Authority did not require payment of any additional fee.
3.	☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically Claims Nos.:
4.	☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by Claims Nos.:
Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest. ☐ No protest accompanied the payment of additional search fees.	

Form PCT/ISA/210 (continuation of first sheet) (1 July 1999)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.

PCT/US 01/26317

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 9845437	A	15-10-1998	AU 6956798 A EP 0973699 A2 JP 2001518793 T WO 9845437 A2	30-10-1998 26-01-2000 16-10-2001 15-10-1998
WO 0058473	A	05-10-2000	AU 3774500 A EP 1165784 A2 WO 0058473 A2	16-10-2000 02-01-2002 05-10-2000
WO 0161622	A	23-08-2001	AU 3259201 A WO 0161622 A1	27-08-2001 23-08-2001

INTERNATIONAL SEARCH REPORT				International Application No.	
Information on patent family members				PCT/US 00/08260	
Patent document cited in search report	Publication date	Patent family member(s)	Publication date		
WO 9906439 A	11-02-1999	AU 8555298 A	22-02-1999		
		EP 0994899 A	25-04-2000		

Form PCTA/2001/10 issued jointly by WIPO and EPO

フロントページの続き

(51)Int.Cl. ⁷	F I	テーマコード(参考)
C 1 2 N 1/15	C 1 2 N 1/15	4 C 0 8 4
C 1 2 N 1/19	C 1 2 N 1/19	4 C 0 8 5
C 1 2 N 1/21	C 1 2 N 1/21	4 H 0 4 5
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 33/53	M
G 0 1 N 33/566	G 0 1 N 33/566	
	C 1 2 N 5/00	A
	A 6 1 K 37/02	

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

(72)発明者 クラスノー、ランディ・イー

アメリカ合衆国カリフォルニア州 9 4 3 0 5 ・ スタンフォード・サンタフェアベニュー 8 1 7

Fターム(参考) 2G045 AA29 AA34 AA35 BB20 BB41 BB46 BB50 CB01 CB02 DA13
DA36 FA12 FA29 FB02 FB03 FB05 FB12 GC15
4B024 AA01 AA12 BA36 BA41 BA44 CA04 DA02 EA02 GA11 HA01
4B063 QA01 QQ43 QR32 QR55 QS33 QS34
4B064 AG27 AG31 CA10 CA19 CC24 DA01 DA14
4B065 AA90X AA99Y AB01 AC14 BA02 CA24 CA25 CA44 CA46
4C084 AA02 BA01 BA02 BA08 BA22 BA23 NA14 ZA81 ZB26
4C085 AA14 BB11 DD61 EE01
4H045 AA10 AA11 AA30 BA10 CA41 DA76 DA86 EA28 EA51 FA74

专利名称(译)	肾特异蛋白		
公开(公告)号	JP2004507242A	公开(公告)日	2004-03-11
申请号	JP2002522268	申请日	2001-08-23
[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	ウォーカーマイケルジー クラスノーランディイー		
发明人	ウォーカー、マイケル・ジー クラスノー、ランディ・イー		
IPC分类号	G01N33/50 A61K38/00 A61K39/395 C07K14/47 C07K14/82 C07K16/32 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	C07K14/47		
FI分类号	C12N15/00.ZNA.A A61K39/395.N A61K39/395.T C07K14/82 C07K16/32 C12N1/15 C12N1/19 C12N1/21 C12P21/02.C C12P21/08 C12Q1/68.A G01N33/15.Z G01N33/50.Z G01N33/53.D G01N33/53.M G01N33/566 C12N5/00.A A61K37/02		
F-TERM分类号	2G045/AA29 2G045/AA34 2G045/AA35 2G045/BB20 2G045/BB41 2G045/BB46 2G045/BB50 2G045/CB01 2G045/CB02 2G045/DA13 2G045/DA36 2G045/FA12 2G045/FA29 2G045/FB02 2G045/FB03 2G045/FB05 2G045/FB12 2G045/GC15 4B024/AA01 4B024/AA12 4B024/BA36 4B024/BA41 4B024/BA44 4B024/CA04 4B024/DA02 4B024/EA02 4B024/GA11 4B024/HA01 4B063/QA01 4B063/QQ43 4B063/QR32 4B063/QR55 4B063/QS33 4B063/QS34 4B064/AG27 4B064/AG31 4B064/CA10 4B064/CA19 4B064/CC24 4B064/DA01 4B064/DA14 4B065/AA90X 4B065/AA99Y 4B065/AB01 4B065/AC14 4B065/BA02 4B065/CA24 4B065/CA25 4B065/CA44 4B065/CA46 4C084/AA02 4C084/BA01 4C084/BA02 4C084/BA08 4C084/BA22 4C084/BA23 4C084/NA14 4C084/ZA81 4C084/ZB26 4C085/AA14 4C085/BB11 4C085/DD61 4C085/EE01 4H045/AA10 4H045/AA11 4H045/AA30 4H045/BA10 4H045/CA41 4H045/DA76 4H045/DA86 4H045/EA28 4H045/EA51 4H045/FA74		
优先权	09/645961 2000-08-24 US		
外部链接	Espacenet		

摘要(译)

本发明提供了编码肾特异性蛋白质的哺乳动物cDNA。它还提供cDNA，片段，互补物，变体和编码的蛋白质或其部分，用于诊断和治疗肾脏疾病，以及抗体的用途。

	ライブラリ 1	ライブラリ 2	ライブラリ 3	...	ライブラリ N
遺伝子 A	1	1	0	...	0
遺伝子 B	1	0	1	...	0