

(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2004-516813

(P2004-516813A)

(43) 公表日 平成16年6月10日(2004.6.10)

(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 31/7088	A 6 1 K 31/7088	4 B O 2 4
A 6 1 K 38/00	A 6 1 K 39/395 E	4 B O 6 3
A 6 1 K 39/395	A 6 1 K 39/395 T	4 B O 6 4
A 6 1 K 45/00	A 6 1 K 45/00	4 B O 6 5
	審査請求 未請求 予備審査請求 有 (全 112 頁) 最終頁に続く	

(21) 出願番号 特願2002-508012 (P2002-508012)
 (86) (22) 出願日 平成13年6月29日 (2001.6.29)
 (85) 翻訳文提出日 平成14年12月26日 (2002.12.26)
 (86) 国際出願番号 PCT/US2001/020729
 (87) 国際公開番号 W02002/002772
 (87) 国際公開日 平成14年1月10日 (2002.1.10)
 (31) 優先権主張番号 60/215,454
 (32) 優先日 平成12年6月30日 (2000.6.30)
 (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 ラセク、アミー・ケイ・ダブリュ
 アメリカ合衆国カリフォルニア州9461
 1・オークランド・#5・フォーティファ
 ーストストリート 237

最終頁に続く

(54) 【発明の名称】 関連腫瘍マーカー

(57) 【要約】

本発明は、ECM関連タンパク質をコードするcDNAを提供する。また疾患、特に結腸癌と肺癌の診断、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。さらに本発明は、タンパク質と遺伝形質転換モデル系の生成のための発現ベクターと宿主細胞を提供する。

```

10      19      28      37      46      55
5' ATG CTA TCG AGA GTC TTC CTC CAA ACA TCT TCC GAT TCG TTC CTT TAA CCC ATC
...
64      73      82      91      100     109
TAG ATC TTC GTG GAA ATC AAT TAC AAA CAT TGC CTT ATG TTG GTT TTC TCG AAC
...
118     127     136     145     154     163
ACA TTG GTC GAA TAT TCG ATC TTC AGT TCG AGG ACA ACA AAT GGG CCT GCA ATT
...
172     181     190     199     208     217
GTG ACT TAT TGC AGT TAA AAA CTT GGT TGG AGA ACA TGC CTC CAC AGT CTA TTA
...
226     235     244     253     262     271
CTT GGT GAT GTT CTC TCC NAC AGC CCT CCA TTT TTT AAA GGA AGT ATA CTC AGT
...
280     289     298     307     316     325
AGA CTA AAG AAG GAA TCT ATT TGC CCT ACT CCA CCA CTG TAT GAA GAA CAT GAG
...
334     343     352     361     370     379
GAT CCT TCA GGA TCA TTA GAT CTG GCA GCA ACA TCT TCA ATA AAC GAT AGT GGC
    
```

【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) - c) から選択される。

a) SEQ ID NO : 2 及びその相補体

b) SEQ ID NO : 3 - 13 またはその相補体から選択した SEQ ID NO : 2 の断片、そして

c) SEQ ID NO : 14 - 15 またはその相補体から選択した SEQ ID NO : 2 の変異体。 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、及び 20

b) 宿主細胞株からタンパク質を回収する方法を含む。

【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。 30

【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。

【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であって、 50

c) SEQ ID NO: 1の生物学的活性な部分。

【請求項14】

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。

10

【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、

20

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。

30

【請求項20】

結腸と肺の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】

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癌の治療方法。

【請求項23】

癌が結腸癌であることを特徴とする請求項22に記載の方法。

40

【請求項24】

癌が肺癌であることを特徴とする請求項22に記載の方法。

【発明の詳細な説明】

【0001】

この出願は2000年1月30日に提出した米国特許仮出願第60/215,454号の利益を請求するものであり、すべての特許出願は言及することをもって本明細書の一部とする。

【0002】

(技術分野)

本発明は細胞外マトリックス(ECM)関連腫瘍マーカーをコードするcDNA及びcD

50

NAの使用そして腫瘍、特に結腸と肺の癌の診断と治療においてコード化されたタンパク質に関連する。

【0003】

(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長所があり、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の核酸とタンパク質配列の比較は、ヒト配列の調査を加速する。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。

10

【0004】

癌と悪性腫瘍は連続した細胞増殖と細胞死によって特徴付けられ、遺伝学と環境の両方に因果関連がある。癌マーカーは、癌への家族性素因の決定また早期検診そして様々な癌の予後において非常に重要である。

【0005】

結腸直腸癌は4番目に発生率の高い癌であり、米国では癌による死の2番目に高い原因となっており、毎年約13万人が新患となり、5万5千人が死亡する。結腸と直腸癌は多くの環境危険因子を共有しており、両者とも個々に特有の遺伝的症候群が見られる。(結腸直腸癌の概説にはPotter, J D (1999) J Natl Cancer Institute 91:916-932を参照)。結腸癌は男女でおよそ等しい頻度で発生する唯一の癌である。また米国では結腸癌の診断5年後の生存率は約55%である(Riesら(1990) National Institutes of Health, DHHS Publ No. (NIH) 90-2789)。

20

【0006】

結腸癌は遺伝子と環境の両者と因果関係がある。幾つかの分子経路は結腸癌の発生と関係があり、これらの経路すべての鍵遺伝子の発現は先天性または後天性突然変異或いは過剰メチル化により失われる可能性がある。発現の変化が結腸癌の早期指標あるいは結腸癌の発生の素因を提供し、疾患の治療に役立つ可能性のある治療標的をも提供する可能性がある。遺伝子を同定する特別な必要性がある。

30

【0007】

結腸癌の素因、発生、進行と関連のある多数の遺伝子が同定されてきた。例えば、ヒトの腫瘍ではDNAメチル化の異常パターンが常に起こることが十分に知られている。特に結腸癌では、結腸癌に先行する悪性ポリープなど腫瘍進行においてこれらの変化が早期に発生することが見つけられている。DNAメチル化を行う酵素であるDNAメチルトランスフェラーゼは、結腸癌または癌に先行する良性ポリープの患者の組織学的に正常な粘膜で著しく増加する。そしてこの増加は結腸腫瘍の進行中継続する(Wafikら(1991) Proc Natl Acad Sci USA 88:3470-3474)。家族性大腸腺腫症(FAP)は結腸癌に先行するまれな常染色体優性症候群であり、大腸腺腫様ポリーポーシス(APC)遺伝子の先天性突然変異によって起こる。APC遺伝子は、APC - カテニン - Tcf (T-細胞因子)経路の一部である。この経路の障害により、結腸上皮細胞の規則的な複製化、接着性、移動性が失われ、それがポリープの成長につながる。遺伝性非腺腫性結腸癌(HNPCC)は別の先天性常染色体優性症候群であり、結腸癌の早期発症および他の癌の発生の性向とは区別される。HNPCCは、DNAミスマッチ修復(MMR)経路の1つか複数の遺伝子の突然変異から発生する。2つのヒトMMR遺伝子MSH2およびMLH1の突然変異は、これまでに同定されたHNPCCファミリーの大部分に見出される。ほとんどすべての結腸癌は、エストロゲン受容体(ER)遺伝子が不活性の細胞から発生する。ER遺伝子転写の不活性化は年齢に関係しており、ER遺伝子の過剰メチル化に関連する(Issara(1994) Nature Genetics 7:536-540)。培養した結腸癌細胞に外来性ER遺伝子を導入すること

40

50

により、著しく成長が抑制される。

【0008】

結腸癌、癌発生の早期指標を提供する可能性のある疾患の発生と進行と関連する多数の遺伝子変更が存在することは明らかであり、疾患進行のモニタ - または治療標的の提供にも使用される可能性がある。

【0009】

細胞外マトリックス (ECM) は、細胞から分泌された糖タンパク質、多糖、プロテオグリカンおよび他の高分子の複雑なネットワークである。また細胞に接着性、移動性、シグナル伝達のために機械的な足場を提供する。腫瘍細胞は転移性になるプロセスにおいて細胞ECM固着の必要性およびECM仲介成長制御をしばしば克服するようになる (Ruoslahti, E. (1996) Sci. Am. 275: 72-77)。このようにしてECMタンパク質の変異は、ある種の癌に対するマーカーの追加供給源と可能性のある治療標的を提供する。多くのECMタンパク質は膜貫通ドメインの存在によって特徴付けられる。そしてロイシンリッチリピート (LLR) などのタンパク質 - タンパク質または細胞 - マトリックス相互作用ドメインはシグナル伝達、タンパク質 - タンパク質相互作用における細胞接着と関連があると思われる。

10

【0010】

ECM関連タンパク質をコードするcDNAの発見は、癌、特に結腸と肺の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を満たす。

【0011】

20

(発明の概要)

本発明はECM関連タンパク質をコードするcDNA、ECMRPの発見に基づいており、癌、特に結腸と肺の癌の診断と治療に有用である。

【0012】

本発明は、SEQ ID NO: 1のアミノ酸配列を有するタンパク質をコードする核酸配列を含む単離したcDNAを提供する。本発明はまた、SEQ ID NO: 2の核酸配列を含む群から選択した単離したcDNAまたはその相補配列、SEQ ID NO: 3 - 13から選択したSEQ ID NO: 2の断片、SEQ ID NO: 14 - 15から選択したSEQ ID NO: 2の変異体を提供する。更に本発明は、ECM関連タンパク質をコードするcDNAあるいは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、cDNAを含むベクター、ベクターを含む宿主細胞およびECM関連タンパク質を作製するためにcDNAを使用する方法を提供する。更に本発明は、ECM関連タンパク質をコードするcDNAを含むベクターから成る遺伝形質転換細胞株または生物を提供する。本発明はまた、SEQ ID NO: 2 - 15から成る群から選択したcDNAの断片、変異体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、変異体またはその相補配列のうちの少なくとも1つを含む基質を提供する。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、精製方法で使用することが可能なcDNAまたはその相補配列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

【0013】

40

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癌、特に結腸と肺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また断片あるいは変異体またはその相補配列はアレイのエレメントを含み得る。

【0014】

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また断片あるい

50

は変異体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 c D N A を分子または化合物に結合する、そして c D N A への特異結合を検出する、それによって c D N A 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 D N A 分子、R N A 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。

【 0 0 1 5 】

本発明は、S E Q I D N O : 1 のアミノ酸配列から成る群から選択したタンパク質または精製されたタンパク質、S E Q I D N O : 1 のアミノ酸配列と少なくとも 8 5 % の相同性を有する変異体、S E Q I D N O : 1 の抗原性エピトープまた S E Q I D N O : 1 の生物学的活性な部分を提供する。本発明は精製されたタンパク質と医薬用担体から成る組成物も提供する。本発明はさらに、あるタンパク質を使って、c D N A 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方法を提供し、その方法の内には、特異結合を可能にし特異結合を検出できる条件下で同タンパク質を分子または化合物を混合すること、それによって同タンパク質に特異結合するリガンドを同定ことが含まれる。ある実施様態においては、それらの分子または化合物は D N A 分子、R N A 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。別の実施様態において、リガンドは癌、特に結腸と肺癌を患う患者を治療するために使用される。

10

【 0 0 1 6 】

本発明は、あるタンパク質を使って、そのタンパク質に特異的に結合する抗体で、その抗体の存在または量が癌、特に結腸と肺癌の診断となるような抗体あり、そのような抗体を見つけるためにある被検サンプルをスクリーンする方法を提供し、その方法は同被検サンプルから抗体を単離し、特異結合を可能にする条件下でその単離した抗体を同タンパク質に接触させ、その結合したタンパク質から抗体を解離し、知られる標準でその抗体の量を比較することが含まれる。

20

【 0 0 1 7 】

さらに本発明は、あるタンパク質を使って抗体を準備し精製する方法を提供し、その方法には、抗体反応を誘発する条件下で同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、ある基質に同タンパク質を付着させること、同タンパク質への特異結合を可能にする条件下で同基質を単離した抗体と接触させること、同タンパク質から同抗体を解離すること、そしてそれによって精製した抗体を得ることが含まれる。

30

【 0 0 1 8 】

本発明は、癌特に結腸と肺の癌で発現するタンパク質に特異結合する精製した抗体を提供する。本発明はまた、ある抗体を使って癌、特に結腸と肺の癌を診断する方法を提供し、その方法には、結合した抗体の量を知られる標準と比較し、それによって癌、特に結腸と肺の癌の存在を確定することが含まれる。さらに本発明は癌、特に結腸と肺の癌を治療するためにある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、精製されたその抗体と医薬用担体から成る組成物を投与することが含まれる。

【 0 0 1 9 】

本発明は、内在性ポリヌクレオチドの発現を破壊するために異種マーカ―遺伝子を哺乳動物のゲノム D N A に挿入する方法を提供する。さらに本発明は、ある c D N A を使用して哺乳動物モデル系を作製する方法を提供し、その方法には、S E Q I D N O : 2 - 1 5 から選択された c D N A を含むあるベクターを作製すること、そのベクターを E S 細胞に形質転換すること、形質転換した E S 細胞を選択すること、形質転換した E S 細胞を哺乳動物胚盤胞に微量注入すること、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すること、前記メスが生殖細胞系で c D N A を含むキメラ子孫を産むこと、そしてホモ接合性系、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ことが含まれる。

40

【 0 0 2 0 】

50

(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。

【0021】

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。

10

【0022】

定義

「ECM関連タンパク質」は、天然、合成、半合成或いは組換え体など全ての種(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哺乳動物)から得られる精製されたタンパク質を指す。

20

【0023】

「アレイ」は、基質上の少なくとも2つのcDNAあるいは抗体の規則正しい配列を指す。cDNAあるいは抗体の少なくとも1つが調節または標準を表す。そして他方が目的の診断または薬剤のcDNAあるいは抗体を表す。基質上の2 - 約40,000のcDNAあるいは2 - 約40,000のモノクローナルまたはポリクローナル抗体の構成により、各cDNAと少なくとも1つの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、あるいは各抗体と抗体が特異結合する少なくとも1つのタンパク質抗体タンパク質複合体は、確実に個別に区別できる。

【0024】

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。

30

【0025】

「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3'または5'配列を表す、そしてイントロンが欠けている。

【0026】

「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

【0027】

「組成物」とは、ポリヌクレオチドと標識成分、精製されたタンパク質と医薬用担体、抗体と標識成分などを指す。

【0028】

「誘導體」とは化学修飾されたcDNAあるいはタンパク質を指す。cDNAの誘導體化にはクエオシン(queosine)あるいはヒポキサンチンなどの類似体等非従来型塩

50

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。

【0029】

「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、あるいは減少、また下方調節発現を指す。

【0030】

「疾患」とは、cDNAとECM関連タンパク質が他と異なって発現する症状、疾病、または症候群を指す。そのような疾患には癌、特に結腸と肺の癌が含まれる。 10

【0031】

「断片」とは長さが約50から約4000塩基対の連続したヌクレオチドを指す。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。このようなりガンドは、複製、転写または翻訳を調節する治療として有用である。

【0032】

「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ハイブリダイゼーションの条件、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。 20

【0033】

「標識成分」とは、cDNAがタンパク質に結合接着するあるいは組み込まれることが可能な任意の可視または放射性標識を指す。可視標識には、限定されるものではないが、アントシアニン、緑色蛍光タンパク質(GFP)、グルクロニダーゼ、ルシフェラーゼ、Cy3とCy5などが含まれる。放射性マーカーには、水素、ヨウ素、亜リン酸、硫黄などの放射性状態が含まれる。

【0034】

「リガンド」とは、ポリヌクレオチドあるいはタンパク質の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。このようなり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する。そしてミネラル、補助因子、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ている可能性がある。 30

【0035】

「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使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。

【0036】

「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。 40

【0037】

「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を指す。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。

【0038】

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。

【0039】

「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 50

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。

【0040】

「タンパク質」とは、ポリペプチドあるいはその任意の部分指す。タンパク質の「部分」とは、少なくとも1つの生物学的活性を保持するであろうアミノ酸配列の長さ、PFAMまたはPRINTS分析によって同定されるドメインあるいはPROTEAN プログラム(DNA STAR, Madison WI)のKyte-Doolittle 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。

10

「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。

【0041】

「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。

【0042】

配列に適用される「類似性」とは、Smith-Waterman アルゴリズム(Smith及びWaterman (1981) J Mol Biol 147:195-197)あるいはBLAST2 (Altschulら (1997) Nucleic Acids Res 25:3389-3402)などの標準化されたアルゴリズムを用いてアラインメントされた少なくとも2つの配列間で一致する分子あるいは残基の定量化(通常は%)を指す。BLAST2は、アラインメントを最適化するために配列の1つでギャップを挿入するまた2つの配列をより有意に比較できる標準化された再現性のある方法で使用される可能性がある。タンパク質では特に、保存的置換では類似性が同一性よりも大きい。例えば、ロイシンまたはイソロイシンのバリンは報告された割合を計算する時にカウントされる。保存と考えられた置換が当分野で公知である。

20

【0043】

「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、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。

30

【0044】

「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。

【0045】

「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、BLAST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に対して高一一致率を有し、100塩基に付き約3塩基が異なる。「1塩基ヌクレオチド多形性」(SNP)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。

40

【0046】

発明

本発明は、結腸と肺癌の性質決定、診断、治療におけるECM関連タンパク質をコードす

50

る cDNA の発見、cDNA またはその断片、タンパク質またはその部分、直接あるいは組成物としての使用に基づくものである。

【0047】

本発明の ECM 関連タンパク質をコードする核酸は、ヌクレオチドそして/またはアミノ酸配列アラインメント用のコンピュータ検索を利用して最初に Incyte クローン 2743093 で同定した。SEQ ID NO: 2 は下記の重複そして/あるいは伸長した核酸配列に由来する。(SEQ ID NO: 3 - 13): Incyte クローン 2743093H1 (SKINDIA01)、4876623F9 (COLDN01)、2316239T6 (OVARNOT02)、6258015F8 (BMARTXT06) およびショットガン配列 7677606J1、71111915V1、71112850V1、71262960V1、71264035V1、71113484V1、71114738V1。

【0048】

一実施態様では、本発明は図 1A から 1J で表示されるようにアミノ酸配列 SEQ ID NO: 1 を含むポリペプチドを提供する。ECM 関連タンパク質は長さが 546 アミノ酸であり、N114 と N446 で 2 つの潜在的 N-グリコシル化部位、S59、T215、T244、S305、S381、S425、S453、T521 で 8 つの潜在的カゼインキナーゼ I リン酸化部位、S2 と T375 で 2 つの潜在的プロテインキナーゼ C リン酸化部位および Y510 で 1 つの潜在的チロシンキナーゼリン酸化部位を有する。PFAM 分析は、SEQ ID NO: 1 のアミノ酸残基 T93 と P188 の間でタンパク質タンパク質結合作用と関連する幾つかのロイシンリッチリピートドメイン (LRR) を示す。さらに PFAM 分析は残基 N222 から P272 の C 末端 LRR を示す。さらに PRINTS 分析は LRR の存在を L118 から L131 まで、また SEQ ID NO: 1 の M163 から L176 まで確認する。さらに HMMR 分析は膜貫通ドメインの存在を SEQ ID NO: 1 の残基 I316 と V335 の間で示す。有用な抗原性エピトープは ECMRP の約 L60 から約 L175 まで伸長する、そして ECMRP の生物学的活性な部分は約 L118 から約 L131 までまた約 M163 から約 L176 まで伸長する。ECM 関連タンパク質を特異結合する抗体は、癌、特に結腸と肺癌を同定する診断アッセイにおいて有用である。

【0049】

図 2 は、TAQMAN 分析により ECMRP 発現を分析した多様な正常な成熟した組織の結果を示す。ECMRP の著しい発現は、肝臓、前立腺、結腸、甲状腺、下垂体、副腎組織にのみ見られる。また心臓、脳、肺、骨格筋、腎臓、膵臓、脾臓、胸腺、卵巣、小腸、末梢血白血球では検出されない。

【0050】

表 1 は、LifeSeq データベース (Incyte Genomics) における cDNA ライブラリのノーザン分析による組織のカテゴリーの ECMRP の発現を示す。結果は、ECMRP の発現が消化器系、内外分泌系、女性生殖器 (前立腺など)、呼吸器系、皮膚など多様な組織のカテゴリーに分布していることを示す。上記、表 1 と図 2 の結果の間で観察される差異は、恐らく LifeSeq データベースの cDNA ライブラリの胎児の病変組織の高発生率を反映する。

【0051】

図 3 は、TAQMAN 分析 (Applied Biosystems) による正常な結腸組織と比較した結腸癌組織サンプルの ECMRP の発現、また正常な肺と比較した肺腫瘍サンプルを示す。結果は、調べたサンプル 9 つのうち 6 つで結腸腫瘍における ECMRP の発現の増大を示す (提供者 ID: 3581、3583、3647、3649、3479、4614)。癌と正常な組織の間で発現の差異が少なくとも 1.2 倍であることが観察されるならば結果は重大であると考えられた。

【0052】

表 2 は、正常な結腸組織に対する結腸癌または結腸ポリープ組織における、また正常な肺

に対する肺腫瘍組織におけるECMRPの発現を比較するマイクロアレイ分析の結果を示す。結果は、調べた患者14人中7人で結腸腫瘍またはポリープにおけるECMRPの発現の増大を示す(提供者ID: 4614、3755、3311、3754、3583、3839、3581)。ECMRPの発現の増大は調べた肺腫瘍のサンプル10のうち2つでも観察された(提供者ID: 5796、5800)。癌と正常な組織の間で発現の差異が少なくとも1.5倍であることが観察されるならば、示差発現(列2)は重大であると考えられた。

【0053】

正常な結腸組織と比較した場合、ECMRPの発現の増大はヒトの結腸直腸腺癌細胞株、HT-29でも観察された。(データ表示されず)

ECM関連タンパク質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BLAST2とZOOSEQデータベース(Incyte Genomics)を使用して同定された。これらの好適な変異体は、以下の表で示されるように約87%から89%の同一性を有する。列1、変異cDNAのSEQ ID var、列3、変異cDNAのクローン番号、列4、ヒトのcDNAへの一致率、列5、ヒトのcDNAへの変異cDNAのアラインメント。

SEQ ID _{var}	cDNA _{var}	種	一致	Nt _H アラインメント
14	703545995J1	イヌ	89%	2-192
15	112357_Mm.1	サル	87%	3003-3173

10

20

【0054】

当業者であれば、遺伝暗号の縮重の結果、ECM関連タンパク質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。これらの組合せは、天然のECM関連タンパク質をコードするポリヌクレオチドに適用されるような標準トリプレット遺伝暗号を基に作られるものであり、このような変異は全て明確に開示されているものと考えられる。

【0055】

SEQ ID NO: 2-15のcDNA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、SEQ ID NO: 14-15は、ヒトの結腸または肺癌のモデル系である遺伝形質転換細胞株または生物を産生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。

30

【0056】

cDNA sとタンパク質、断片とその部分の同定と性質決定は米国特許仮出願60/215,454に記載されており、すべての文献は言及することをもって本明細書の一部とする。

40

【0057】

本発明の性質決定と使用

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する。Incyte cDNAを、実施例に記載されているように準備した哺乳動物cDNAライブラリから単離した。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 app

50

lication (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る。5'と3'配列の検証後、ECM関連タンパク質をコードする少なくとも1つの代表的なcDNAが試薬を指定する。

【0058】

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T7DNA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。

10

20

【0059】

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関係する。これらの断片は、元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cDNAsの側面に位置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配列あるいは除去配列は、不完全に構築された配列を完全な配列に組織化することにより、それぞれ除去、回復が可能である。

30

【0060】

核酸の伸長

本発明の配列は、当分野で周知の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。

40

50

【0061】

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、設計することは可能である。そしてECM関連タンパク質、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: 2-15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。

10

【0062】

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。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, Plainview NYに概説されている。

20

30

【0063】

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またはcDNA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、突然変異及び一塩基多型性を同定するためにハイブリダイゼーションプローブあるいは標的として使用され得る。アレイを用いることで、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）USPN 5,474,796、Schenarら（1996）Proc Natl Acad Sci 93:10614-10619、Hellerら（1997）Proc Natl Acad Sci 94:2150-2155、Hellerら（1997）USPN 5,605,662等を参照）。

40

【0064】

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。ハイブリダイゼーションプローブは、特定の染色体、染色体の特定領域または人工形成の染色体にハイブリダイズされる。そのような産物には、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、或いは単一染色体から作成されるライブラリのcDNAが含まれる。

50

【0065】

発現

ECM関連タンパク質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、新規な制限部位の生成、グリコシル化パターンの変更、コドン優先の変更のために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ソースからの転写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。

10

【0066】

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領域に結反させた後に、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。

20

【0067】

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、PBLUESCRIPTベクター(Stratagene, La Jolla CA)またはpSPORT1プラスミド(Life Technologies)などの多機能のベクターを用いることができる。これらのベク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のための比色スクリーニング法が可能となる。更にこれらのベクターは、クローニングされた配列における*in vitro*転写、ジデオキシのシーケンシング、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。

30

【0068】

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-または可視マーカ-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-、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。選択的培地上の生存あるいは可視マーカ-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-は、導入された遺伝子により発現したタンパク質の量を推定するためにも用いられる。宿主細胞が望ましい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。

40

【0069】

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(例

50

えはCHO、HeLa、MDCK、MEK293、WI38等)は、American Type Culture Collection (ATCC, Manassas VA) から入手可能であり、外来タンパク質の正しい修飾及び処理を確実にするように選択し得る。

【0070】

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。

10

【0071】

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、バッチワイズ(batchwise)或いは-アミノ-と側鎖-保護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る連続流れ処理で実施され得る。メチルアミン-derivatized 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ポリ(スチレン-co-ジビニルベンゼン)に結合接着する。アミノ酸残基はacid labile Boc (t-ブチル酸素カルボニル)または塩基-labile Fmoc (9-フルオレニルメトキシカルボニル)によりN-保護される。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ノに結合する。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カルボキシル末端とリンカー群の間で切断する。(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。タンパク質あるいはその部分は、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)。

20

30

【0072】

抗体の調製とスクリーニング

限定されるものではないが、ヤギ、ウサギ、ラット、マウス、ヒト細胞株を含む種々の宿主が、ECM関連タンパク質またはその任意の部分の注入によって免疫化され得る。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。

40

【0073】

50

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)。

【0074】

別法では、当分野で周知の方法を用いて、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)。

10

【0075】

ECM関連タンパク質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イムノアッセイ (two-site, monoclonal-based immunoassay) が好ましいが、競合の結合アッセイも用いることもできる (Pound (1998) *Immunochemical Protocols*, Humana Press, Totowa NJ)。

20

【0076】

ラジオイムノアッセイ技術と共にScatchard分析などの様々な方法を用いて、ECMRPに対する抗体の親和性を評価する。親和性を結合定数Kaで表すが、このKaは、平衡状態の下でHSPDE10A抗体複合体のモル濃度を遊離抗体と遊離抗原のモル濃度で除して得られる値である。多数のエピトープに対して親和性が不均一なポリクローナル抗体医薬のKaは、ECMRPに対する抗体の平均親和性または結合活性を表す。特定のエピトープに単一特異的なモノクローナル抗体試薬のKaは、親和性の真の測定値を表す。Ka値が10⁹ ~ 10¹² L/molの高親和性抗体試薬は、タンパク質抗体複合体が激しい操作に耐えなければならないイムノアッセイに用いるのが好ましい。Ka値が10⁶ ~ 10⁷ L/molの低親和性抗体試薬は、ECMRPが抗体から最終的に活性化状態で解離する必要がある免疫精製 (immunopurification) 及び類似の処理に用いるのが好ましい (Catty, D. (1988) *Antibodies, Volume I: A Practical Approach*, IRL Press, Washington DC; Liddell及びCryer (1991) *A Practical Guide to Monoclonal Antibodies*, John Wiley, New York NY)。

30

40

【0077】

アッセイのための分子の標識化

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、³²P-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³⁵S-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

50

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。

【0078】

(診断)

核酸アッセイ

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、疾患の診断のために差次的な遺伝子の発現を検出し定量するために用いられる。同様に、ECM関連タンパク質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疾患には、癌、特に結腸と肺の癌が含まれる。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る可能性がある。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。

10

【0079】

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。

20

【0080】

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、cDNAと動物或いはヒトの何れかの正常な被験者から抽出されたサンプルを結合させることにより達成され得る。精製された配列を既知量用いて行った実験から得た値を正常な対象から得た値と比較することにより、標準ハイブリダイゼーション複合体を定量することができる。このようにして得た標準値は、特定の症状、疾患または障害と診断される患者から得たサンプルから得た値と比較することができる。特定の疾患と関連のある値への標準値からの偏差を、その疾患の診断に用いる。

30

【0081】

このようなアッセイを、動物実験または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ペースで繰り返して、患者の発現レベルが正常な被検者に観察されるレベルに近づき始めたかどうかを測定する。連続アッセイから得られた結果を用いて、数日から数年の期間にわたる治療の効果を示し得る。

【0082】

タンパク質アッセイ

標識されたアミノ酸または特異的ポリクローナル抗体または特異的モノクローナル抗体を用いるタンパク質の検出と定量は、当分野で公知である。このような技術の例としては、二次元ポリアクリルアミドゲル電気泳動法、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)などが挙げられる。これらのアッセイと精製、標識化基準への定量化は当分野で公知である(前出の Ausubel unit 10.1 - 10.6)。2つの非干渉エピトープに反応するモノクローナル抗体を用いた、2部位のモノクローナルベースイムノアッセイ(two-site, monoclonal-based immunoassay)が好ましいが、競合の結合アッセイも用いることもできる。(Coliganら(1997) Current Protocols in Immunology, Wiley-Intersc

40

50

ience, New York NY、前出 Pound を参照)。

【0083】

治療

ECM 関連タンパク質 (SEQ ID NO: 1) と他の ECM 関連タンパク質の領域の間に、特に膜貫通ドメインと幾つかの LLR ドメインの化学的及び構造的類似性が存在する。さらに、表 2 と図 3 で示されているように示差発現が結腸と肺の癌に強く関連している。ECM 関連タンパク質は、癌、特に結腸と肺の癌に明らかに関与している。

【0084】

結腸または肺の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。

10

【0085】

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 of 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。

20

【0086】

核酸を用いる遺伝子発現の変更

ECM 関連タンパク質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配列やアンチセンス分子 (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。転写開始を阻害するように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 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

【0087】

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

【0088】

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

50

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)、ワイプトシン(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。

【0089】

スクリーニング及び精製アッセイ

ECM関連タンパク質をコードする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、内在性遺伝子の活性、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ライブラリまたは複数の分子が複合物とポリヌクレオチドを結合すること、そして一本鎖または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。

10

【0090】

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、ゲル阻止アッセイ(USPN 6,010,849)または網状赤血球可溶化液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抽出する核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に対して回復、育成する抗体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、ゲル阻止アッセイにスーパーシフトが起きる。

20

【0091】

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。次にサンプルが除かれ、cDNAと反応あるいは結合する。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。

30

【0092】

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

【0093】

好適な実施例では、ECM関連タンパク質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る可能性がある。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、溶液中で遊離する、非生物的物质または生物的物质に固定させる(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、細胞表面にペプチドを発現そして位置付けしてきた、組換え核酸で安定的に形質転換された生存または固定している原核宿主細胞をスクリーニングアッセイで使用することが可能である。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化合物の形成を測定することは可能である。スクリーニングされた特定の種類の分子または化合物に応じて、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

40

50

ク質を特異結合する。

【0094】

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

10

【0095】

薬理学

医薬品成分には、所望の所定目的を達成する有効量の活性成分と医薬用担体が含まれる。有効投与量の決定は、当業者の能力の範囲内で行う。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。

【0096】

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。

20

【0097】

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、多くの感染症、癌、薬剤および毒性研究が、低費用、利便性、寿命、生殖可能性、豊富な参照文献ゆえにラットまたはマウス等の齧歯類上で実施される。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調査、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。

30

【0098】

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上で実施される。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により、毒性プロファイルが生成され、薬剤への暴露の後にヒトの健康に及ぼす潜在的な結果を評価する。

40

【0099】

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

50

【0100】

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者に薬剤を1回投与することに基づいている。3回の試験が実施される。：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。

【0101】

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、様々なファミリーの種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するという多くの証拠がある。

【0102】

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。

10

【0103】

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。実験中、最初また間隔を置いて動物は調べられ、モニターされる。

【0104】

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交系であり、ヒト疾患をモデルにするのに用いられ、または治療薬や毒性薬剤を試験する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する可能性がある。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。

20

【0105】

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細胞が担体胚内部に置かれる時、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。

30

【0106】

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。

40

【0107】

ノックアウト分析

遺伝子ノックアウト分析では、哺乳動物遺伝子の領域はネオマイシンホスホトランスフェラーゼ遺伝子(neo:Capecchi(1989)Science 244:1288-1292)等の非哺乳動物遺伝子を含むよう酵素処理で修飾される。修飾された遺伝子は、培養したES細胞に形質転換され、相同組換えにより内因性ゲノムに組み込まれる。挿入された配列は、内在性遺伝子の転写と翻訳を破壊する。形質転換細胞を齧歯類胚に注入し、その胞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

50

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。

【0108】

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。形質転換細胞を胞胚に注入し、胞胚を上記のように移植する。遺伝子組換え子孫または近交系について研究し、可能性のある医薬品を用いて処理し、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。

10

【0109】

非ヒト霊長類モデル

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非ヒト霊長類での治療薬の影響を評価する点で傑出している。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する。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非ヒト霊長類(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、齧歯類のモデルでは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。

20

【0110】

別の実施例では、発展途上の分子生物学技術にタンパク質をコードするcDNAを用いて、限定はされないが、現在知られているトリプレット暗号及び特異的な塩基対相互作用などのcDNAの特性に依存する新しい技術を提供することができる。

【0111】

(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される。また本発明を限定する目的は含まれていない。ヒト結腸ポリブライブラリ、COLDN01の準備が説明される。

30

【0112】

1 cDNAライブラリ作製

COLDN01

COLDN01ライブラリは、16才男子の部分結腸切除術、一時的回腸造瘻術、結腸内視術中に病変下行結腸組織から単離したRNAを用いて作製した。病理学検査では、家族性結腸ポリポシスの凝結での結腸粘膜全体が関係する低グレードの形成異常のある無数の(100以上)腺腫性ポリープが見られた。

40

【0113】

POLYTRON ホモジナイザー(Brinkmann Instruments, Westbury NJ)を使用して、凍結組織をホモジナイズして、グアニジニウムイソチオシアネート溶液に溶解する。溶解産物を、L870M超遠心分離機(Beckman Coulter, Fullerton CA)でSW28ローターを使用して5.7Mの塩化セシウムクッション上で18時間、25,000 rpm、室温で遠心分離した。RNAをpH 4.7の酸性フェノールで抽出した、そして0.3 Mの酢酸ナトリウムと2.5倍量のエタノールを用いて析出して、RNアーゼのない水で再懸濁し、37 でDNアーゼ処理した。pH 4.7の酸性フェノールでの抽出と、酢酸ナトリウムとエタノ

50

ールによる析出を繰り返した。mRNAをOLIGOTEXキット(Qiagen, Chatsworth CA)で単離して、cDNAライブラリの作製のために用いた。

【0114】

mRNAをmRNAのポリ(A)テイルにおいて第一鎖cDNA合成を誘導するよう設計された、NotIプライマーアダプタを含むSUPERSCRIPTプラスミドシステム(Life Technologies)で推奨されたプロトコルに従って処理した。二本鎖cDNAは鈍くなり、EcoRIアダプタに連結反応させ、NotI(New England Biolabs, Beverly MA)で消化した。cDNAをSEPHAROSE CL4Bカラム(APB)上で分画して、そのうち400bpを超えるcDNAはpINCYプラスミド(Incyte Genomics)に連結反応させた。プラスミドpINCYは、引き続いてDH5 大腸菌細胞(Life Technologies)に形質転換させた。

10

【0115】

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)末端を満たすことにより作製した。プラスミドは自己連結反応して、細菌性宿主大腸菌株JM109に形質転換した。

20

【0116】

仲介プラスミド、pSPORT1-RIはEcoRIでは消化しなかった。そしてHindIII(New England Biolabs)で消化し、張出部(overhang)末端をクレノウとdNTPで満たした。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とは異なる。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。

【0117】

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0.8kb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。

30

【0118】

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。キットには、960精製のためリガンドでブロックする96-ウェルから成る。推奨されているプロトコルを下記の変更を除いて使用する。1)細菌をカルベニシリン25mg/lと0.4%のグリセロールで、殺菌したTERRIFIC BROTH(BD Biosciences, Sparks MD)1mlで培養した。2)注射後、細胞を19時間培養し、次いで0.3mlの溶解バッファで溶解した。そして3)イソプロパノール沈殿後、プラスミドDNAペレットを0.1ml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。

40

【0119】

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

50

ム (Applied Biosystems) または MEGABACE 1000 DNA シークエンシングシステム (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 0.25×10^3 濃度の標準 ABI プロトコルとキット (Applied Biosystems) に従って配列化された。別法では、cDNA を APB の溶液と色素を用いて配列化した。

【0120】

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、市販のプライマー分析ソフトウェアを用いて、約 22 個から約 30 個のヌクレオチドの長さで約 50% 以上の GC 含量を有し、かつ約 68 ~ 72 の温度で標的配列にアニールするように設計した。ヘアピン構造及びプライマー-プライマー二量体が生じないようにヌクレオチドを伸長した。

【0121】

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。

【0122】

高忠実度の増幅を USPN 5,932,451 で開示された方法を利用した PCR 法によって得た。PCR 法を DNA ENGINE サーマルサイクラー (MJ Research) を用いて 96 穴プレート内で実施した。反応混合液は、鋳型 DNA 及び 200 nmol の各プライマー、 Mg^{2+} と $(NH_4)_2SO_4$ と 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: 94 C で 15 秒間、ステップ 3: 60 C で 1 分間、ステップ 4: 68 C で 2 分間、ステップ 5: ステップ 2、3、及び 4 を 20 回繰り返す。ステップ 6: 68 C 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: 94 C で 15 秒間、ステップ 3: 57 C で 1 分間、ステップ 4: 68 C で 2 分間、ステップ 5: ステップ 2、3、及び 4 を 20 回繰り返す。ステップ 6: 68 C で 5 分間、ステップ 7: 4 で保管。

【0123】

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 (0.25% reagent in 1X TE, 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。

【0124】

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・シークエンシングのために、消化したヌ

10

20

30

40

50

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 A G A R A C E 酵素 (P r o m e g a) で消化した。伸長させたクローンを T 4 D N A (N e w E n g l a n d B i o l a b s) を用いて p U C 1 8 ベクター (A P B) に再連結し、P f u D N A ポリメラーゼ (S t r a t a g e n e) で処理して制限部位の張出部 (o v e r h a n g) 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 L B / 2 X カルベニシリン培養液の 3 8 4 ウェルプレートに 3 7 で一晩培養した。

【 0 1 2 5 】

細胞を溶解して、T a q D N A ポリメラーゼ (A P B) 及び P f u D N A ポリメラーゼ (S t r a t a g e n e) を用いて以下の手順で D N A を P C R 増幅した。ステップ 1 : 9 4 で 3 分間、ステップ 2 : 9 4 C で 1 5 秒間、ステップ 3 : 6 0 C で 1 分間、ステップ 4 : 7 2 C で 2 分間、ステップ 5 : ステップ 2、3 及び 4 を 2 9 回繰り返す。ステップ 6 : 7 2 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 定量試薬 (M o l e c u l a r P r o b e s) で D N A を定量化した。D N A 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 2 0 % 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 エネルギートランスファー 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。

【 0 1 2 6 】

5 c D N A クローンの相同性検索と推定タンパク質

配列表の c D N A 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 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 (1 9 9 2 , P r o t e i n E n g i n e e r i n g 5 : 3 5 - 5 1) 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 4 9 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) 。

【 0 1 2 7 】

K a r l i n と A l t s c h u l (1 9 9 3 ; P r o c N a t l A c a d S c i 9 0 : 5 8 7 3 - 5 8 7 7) 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 1 0 0 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、厳密な一致のストリンジェンシーを約 4 0 の下限 (不要な塩基のために 1 - 2 % のエラーがある) から約 7 0 の 1 0 0 % 一致まで設定した。

【 0 1 2 8 】

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、ヌクレオチド配列をアラインメントする「b l a s t n」、ヌクレオチド配列またはアミノ酸配列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

10

20

30

40

50

format: 1; Penalty for mismatch: -2; オープンギャップ5 及び伸長ギャップ: 2 ペナルティ; ギャップ x drop-off: 50; Expect: 10; Word Size: 11; 及びFilter: on 同一性は配列全体の長さにより測定する。Brenner ら (1998; Proc Natl Acad Sci 95: 6073 - 6078, 文献は、引用することをもって本明細書の一部とする) は配列同一性によって構造的相同性を同定する能力に関してBLASTを分析した。少なくとも150 残基の配列アラインメントには信頼できる閾値は30%の同一性であり、少なくとも70 残基のアラインメントに関しては40%であることがわかった。

【0129】

10

この出願のcDNAをアセンブルしたコンセンサス配列あるいはLIFESEQ GOLDデータベース(Incyte Genomics)で見つけた鋳型と比較した。cDNA、伸長、全長およびショットガン・シーケンシングプロジェクトの構成エレメント配列をPHRED分析にかけて、質のスコアを割り当てた。質のスコアが許容できるすべての配列を多様な予備処理にかけ、経路を編集する。そして低品質の3' 末端、ベクターとリンカー配列、ポリAテイル、Alu 繰返し、ミトコンドリア及びリボゾーム配列、細菌汚染配列を除去する。編集した配列は少なくとも長さが50 bpであり、ジヌクレオチド反復、Alu反復などの情報に乏しい配列と反復エレメントを"Ns"あるいはマスクしたものに置換した。

【0130】

20

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

【0131】

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

30

40

【0132】

構築に続いて、鋳型をBLAST、モチーフその他の機能的分析にかけて、1997年3月6日に提出したUSSN 08/812, 290、USSN 08/811, 758、1997年10月9日に提出したUSSN 08/947, 845、1998年3月4日に提出したUSSN 09/034, 807等に記載されているタンパク質階層に分類した。3つの全フォワードリーディングフレームで各鋳型を翻訳することにより、またHMMER ソフトウェアパッケージ(Washington University School of Medicine, St. Louis MO; <http://pfam.wustl.edu/>)を用いて隠れマルコフモデル(HMM)を基にした保存されたタン

50

パク質ファミリドメインのPFAMデータベースに対する各翻訳を検索することにより鑄型を分析した。MACDNASIS PROソフトウェア(Hitachi Software Engineering)とLASERGENEソフトウェア(DNASTAR)を用いてcDNAを更に分析した。そして公共のデータベース、例えばGenBankのげっ歯類、哺乳動物、脊椎動物、原核生物、真核生物のデータベースと、SwissProt、BLOCKS、PRINTS、PFAMそしてPrositateに問い合わせた。

【0133】

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

【0134】

7 ハイブリダイゼーション技術と分析

組織サンプルの調製

Huntsman Cancer Institute, (Salt Lake City, UT)が、一致した正常な結腸と癌性の結腸または結腸ポリープ組織のサンプルを提供した。下記の様に記載されている。: 供与者3754、性別および年齢不明、pendunculated 結腸ポリープ; 供与者3755、性別および年齢不明、ポリープ、家族病歴に結腸癌; 供与者3583、58才、男性、腺管絨毛腺腫(ポリープ)、患者は過形成性ポリープとも診断された。; 供与者3311、85才、男性、浸潤、低分化腺癌、転移性、2/9 リンパ節陽性、TNM分類; T4、N1、Mx、患者は多発管状腺腫とも診断された。; 供与者3839、60才、性別不明、結腸癌、病理学報告無し。; 供与者4614、67才、性別不明、結腸腺癌、中分化、DUKE'S B、TNM分類; T3、N0。表2では、供与者サンプル3754、3755、3311を供与者3753と名付けられた、3人の提供者の正常な結腸組織を集めた通常の対照組織と比較した。他のすべての比較を同一の提供者からの一致した正常組織および腫瘍組織に対して行った。

【0135】

一致した正常組織と腫瘍組織のサンプル(提供者ID5796、5800)をRoy Castle International Institute for Cancer Research (Liverpool, United Kingdom)から得た。提供者5796は66才の男性であり、有棘細胞癌を患っている。提供者5800は75才の女性であり、有棘細胞癌を患っている。

【0136】

Clontech Laboratories Inc. (Palo Alto CA)から、下記の規準化されたヒト、成人、正常な組織の第一鎖cDNA調製を得た。心臓、25-59才の16人の白人男女から集めた; 脳(全体)、43-55才の4人の白人男性から集めた; 胎盤、22-35才の10人の白人女性から集めた; 肺、24-32才の2人の白人女性から集めた; 肝臓、35才の白人男性; 骨格筋、20-60才の35人の白人男女から集めた; 腎臓、24-55才の14人の白人男女から集めた; 膵臓、25-59才の20人の白人男女から集めた; 脾臓、24-61才の6人の白人男女から集めた; 胸腺、18-32才の9人の白人男女から集めた; 前立腺、20-58才の20人の白人から集めた; 精巣、14-64才の45人の白人から集めた; 卵巣、17-60才の7人の白人から集めた; 小腸、15-57才の32人の白人男女から集めた; 結腸、17-76才の20人の白人男女から集めた(内部粘膜を含む); 末梢血白血球、18-40才の白人男女から集めた。すべてのサンプルはHIV-I、HIV-II、B型肝炎、梅毒に

対して陰性であった。

HT - 29 ヒトの結腸直腸腺癌細胞株を American Type Culture Collection (ATCC, Manassas VA) から得た。そしてサプライヤー仕様書に基づいて培養した。

【0137】

基質上の cDNA の固定

下記の方法の1つによって cDNA を基質に適用する。ゲル電気泳動法により cDNA の混合を分画し、キャピラリー転移によりナイロン膜に転移する。別法として、cDNA を個々にベクターに連結反応させ、細菌性宿主細胞に挿入してライブラリを形成する。次に下記の方法の1つによって cDNA を基質に配置する。最初の方法では、個々のコロニーを含む細菌細胞をロボット利用して切りとり、ナイロン膜に配置する。膜を選択培地（用いられたベクターに依存してカルベニシリン、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 LB 寒天に置き、37 で16時間インキュベートする。寒天から膜を取り除き、連続して10% SDS 変性溶液（1.5 M NaCl, 0.5 M NaOH）中に結腸側面を置く。そして溶液を中和（1.5 M NaCl, 1 M Tris, pH 8.0）し、2xSSCで各10分2回行う。次に膜をSTRATALINKER UV架橋剤（Stratagene）でUV照射する。

10

【0138】

2番目の方法では、インサートの側面に位置するベクター配列に相補的なプライマーを用いて、cDNA を30サイクルのPCRで細菌ベクターから増幅する。PCR増幅で1~2 ngの初期量の核酸から5 µgより大きい最終量まで増大する。SEPHACRYL-400 ビーズ（APB）を用いて長さが400 bpから約5000 bpまで増幅した核酸を精製する。手であるいはドット/スロット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、UV照射により固定する。USPN 5, 807, 522に記載されている方法を用いて、精製した核酸を、ロボットで配置して、ポリマーコートされたスライドガラス上に固定する。顕微鏡スライドガラス（Corning, Acton MA）を良く洗って0.1%のSDS中で超音波をかけ、4%フッ化水素酸（VWR Scientific Products, West Chester PA）中でエッチングし、95%エタノール中で0.05%アミノプロピルシラン（Sigma Aldrich）を用いてコーティングする、そして110のオープンで硬化させることにより、ポリマーコートされたスライドガラスを準備する。スライドガラスを処理前後で蒸留水中で広範囲にわたって洗浄する。核酸をスライドガラス上に置き、次にSTRATALINKER UV架橋剤（Stratagene）を用いてアレイをUV照射に暴露することにより固定する。アレイを室温において0.2% SDSで洗浄し、蒸留水で3度洗浄する。リン酸緩衝生理食塩水（PBS）（Tropix, Inc., Bedford MA）中の0.2%カゼイン中において60で30分間アレイをインキュベートした後、0.2% SDS及び蒸留水で洗浄することにより、非特異結合部位をブロックする。

20

30

【0139】

TAQMAN分析のためのプローブ調製

製造者のプロトコルに従いTAQMAN分析のためのプローブを調整した。

40

【0140】

膜ハイブリダイゼーションのプローブ調製

配列表のcDNAに由来するハイブリダイゼーションのプローブを使用して、膜系ハイブリダイゼーションでcDNA、mRNAまたはゲノムDNAをスクリーニングする。cDNAを45 µl TEバッファの濃度40-50 ngに希釈してプローブを調製する。5分間100に加熱し、短時間遠心分離して変性させる。次に変性したcDNAをREDIPRIME チューブ（APB）を加え、青色が均一に分布するまで軽く混合して、次に短時間遠心分離した。5 µlの[³²P]dCTPを管に加え、含有物を37で10分インキュベートする。5 µlの0.2 M EDTAを加えて標識化反応を停止する

50

。そしてPROBEQUANT G-50 microcolumn (APB)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、記載した膜系ハイブリダイゼーションで使用する。

【0141】

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

次の方法を使用して、表2に示されているマイクロアレイ分析のプローブ調製を行った。サンプルのmRNA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における配列表のcDNAをスクリーニングする。9 μ l TEバッファで濃度200 ngにmRNAを希釈し、5 μ l 5xバッファ、1 μ l 0.1 M DTT、3 μ l Cy3またはCy5標識化混合、1 μ l RNアーゼ阻害因子、1 μ l 逆転写酵素、5 μ l 1x酵母調節mRNAを加え、GEMbrightキット (Incyte Genomics)を用いてプローブを調製する。非コード酵母ゲノムDNAから*in vitro*転写により酵母調節mRNAを合成する (W. Lei, unpublished)。定量調節として、サンプルmRNAに0.002 ng、0.02 ng、0.2 ng、2 ngでの調節mRNAの1つの設定をそれぞれ1:100, 000、1:10,000、1:1000、1:100 (w/w)の比で逆転写反応混合に希釈する。mRNA示差発現パターンを調べるために、調節mRNAの2番目の設定を1:3, 3:1, 1:10, 10:1, 1:25, and 25:1 (w/w)の比で逆転写反応混合に希釈する。反応混合液を混合して、37で2時間インキュベートする。次に反応混合液を20分間、85でインキュベートする。そして2つの連続するCHROMA SPIN+TE 30カラム (Clontech, Palo Alto CA)を利用してプローブを精製する。精製したプローブは、プローブをDEPC処理水の90 μ lに希釈させ、2 μ l 1mg/mlグリコーゲン、60 μ l 5M酢酸ナトリウム、300 μ l 100%エタノールを加えて、析出させたエタノールである。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、氷上に格納する。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。

【0142】

膜系ハイブリダイゼーション

1% Sarkosylと1x高リン酸バッファ(0.5 M NaCl, 0.1 M Na₂HPO₄, 5 mM EDTA, pH 7)を含むハイブリダイゼーション溶液では、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、16時間、ハイブリダイズする。ハイブリダイゼーション後、1 mM Tris (pH 8.0), 1% Sarkosylで15分間、25で膜を洗浄する。そして1 mM Tris (pH 8.0)で、各回15分間、25で4回実施する。ハイブリダイゼーション化合物を検出するために、XOMAT-ARフィルム (Eastman Kodak, Rochester NY)を膜に一晩70で暴露する。そして現像してから、視覚で確認する。

【0143】

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

次の方法が、表2に示されているマイクロアレイ分析でを使用された。プローブを5分間65に加熱してから、5415Cマイクロ遠心分離で5分間、9400 rpmで遠心分離する (Eppendorf Scientific, Westbury NY)。そして18 μ l isをアレイ表面で等分して、カバーガラスで覆う。アレイは、顕微鏡スライドより僅かに大きい空洞を有する防水チェンバーに移す。チェンバーのコーナーに140 μ lの5xSSCを加えることにより、チェンバー内部を湿度100%に保持する。このアレイを含むチャンバーを、60で約6.5時間インキュベートする。アッセイを1xSSC, 0.1% SDSで45、10分間洗浄して、0.1xSSC、45、

10

20

30

40

50

10 分間 3 回洗浄した後に、乾燥した。

【0144】

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、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する。そしてハイブリダイゼーション複合体形成後シグナルを検出する。シグナル強度が、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、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する。そして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が、明確な結合した蛍光を与える (Shal on WO95/35505)。

【0145】

ハイブリダイゼーション複合体は、Cy3の励起のためには488nm、Cy5の励起のためには632nmでスペクトル線を生じ得る 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。20x 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、20µm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 System s, Bridgewater NJ) に送られる。アレイと光電子増倍管間に設置されたフィルターを用いて、シグナルを分離する。用いる蛍光色素の最大発光の波長は、Cy3では565nm、Cy5では650nmである。酵母調節 mRNA をプローブ混合液に加えて、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。

【0146】

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、GEMTOOL S プログラム (Incyte Genomics) である。

【0147】

溶液ベースのハイブリダイゼーション (TAQMAN)

製造者のプロトコル (Applied Biosystems) に従い、TAQMAN 分析のためのハイブリダイゼーション反応とハイブリダイゼーション複合体の検出を実施する。

【0148】

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入

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で一致はエラーが1%から2%以内の厳密さになり、少なくとも積スコア70で一致は厳密となる。積スコアが8から40の間を示す分子を選択して、類似または関連する分子を通常は同定する。

【0149】

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。LIFES EQデータベースのすべての配列とcDNAライブラリを系、器官/組織、細胞のタイプにより分類した。分類には、心血管系、結合組織、消化器系、胎芽構造、内分泌系、外分泌腺、女性生殖器、男性生殖器、生殖細胞、血液及び免疫系、肝、筋骨格系、神経系、膵臓、呼吸器系、感覚器、皮膚、顎口腔系、非分類性/混合性または尿路がある。各カテゴリーでは、配列が発現したライブラリの数を数えて、そのカテゴリー内のライブラリの総数を示した。非ノーマライズしたライブラリでは、著しい発現はcDNAの示差発現の存在あるいは不存在を反映する可能性がある。

10

【0150】

9 相補的分子

cDNAに相補的な分子、約5 (PNA) から5000 bp (cDNAインサートの相補体) を用いて、遺伝子発現を検出あるいは阻害する。検出については、実施例7に記載されている。プロモーターが結合するのを妨げることにより転写を阻害するため、相補的分子を最も固有な5'配列に結合するよう設計する。そしてオープンリーディングフレームの開始コドンの5'UTR上流のヌクレオチドを含む。相補的分子にはゲノム配列(エンハンサーまたはイントロン等)があり、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、リボソームがタンパク質をコードするmRNAに結合するのを阻害する。

20

【0151】

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、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エレメントが用いられる場合は3ヶ月以上続く。

30

【0152】

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

【0153】

10 ECM関連タンパク質の発現

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

40

【0154】

Spodoptera frugiperda (Sf9) 昆虫細胞を組換え Autographa californica 核多角体病ウイルス (AcMNPV) で感染させ

50

る。相同組換えによって多角体遺伝子を cDNA と置換する。多角体プロモーターにより cDNA の転写が行われる。上述の精製を可能にする 6 x his で、融合タンパク質としてタンパク質を合成する。精製したタンパク質を次の活性で用いて、抗体を作製する。

【0155】

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 ECM 関連タンパク質を精製して、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 (DNA STAR) を用いて ECM 関連タンパク質のアミノ酸配列を解析し、免疫抗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 K L H (Sigma - Aldrich) に結合させて、免疫抗原性を高める。

【0156】

完全フロイントアジュバントにおいてエピトープ - K L M 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。

【0157】

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。ムノアフィニティークラムは、CNBr - 活性化 SEPHAROSE 樹脂 (APB) に抗体を共有結合させることにより形成する。タンパク質を含む培養液をイムノアフィニティークラムに通し、タンパク質を優先的に吸着する条件下 (例えば洗浄剤が存在する高イオン強度緩衝液) でカラムを洗浄する。結合後、そのタンパク質を、抗体とタンパク質との結合を切るために、例えば、pH 2 ~ 3 のバッファー、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。

【0158】

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、特異結合が複合体成形を示す、標識を保持する基質上の任意の位置でアッセイする。そして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。

【0159】

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、リガンドの多数のクローニング部位に挿入して、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する。また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コンピテント酵母EGY48 [p8op-lacZ]細胞を同時形質転換する比率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-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-を最小限の容積の1xTE(pH7.5)で集め、2%ガラクトース(Gal), 1%ラフィノース(Raf)、80mg/ml5-ブromo-4-クロロ-3-インドリル-d-ガラクトピラノシド(X-Gal)で補われるSD/-His/-Leu/-Trp/-Ura培地で再培養する。次いで青色コロニ-の成長のために検査する。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する。またロイシン(-Leu)を含まない培地でコロニ-成長を生成する。相互作用により、X-Galで成長するコロニ-の青色コロニ-を産生するp8op-lacZレポーター作成物からの-d-ガラクトシダーゼの発現を活性化する。

【0160】

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-を単離して、30で1から2日間、SD/-Trp/-Ura液体で成育することにより多様化する。培地のサンプルをSD/-Trp/-Ura培地で培養する、そしてコロニ-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なく、ヒスチジンを含むSDで成育するコロニ-は、pLexAプラスミドを失った。ヒスチジンを必要とするコロニ-をSD/Gal/Raf/X-Gal/-Trp/-Uraで成育する。そして白色コロニ-を単離して、増殖する。タンパク質と物理的に相互作用するタンパク質をコードするcDNApB42AD-cDNAプラスミドを酵母細胞から単離して、特徴付ける。

【0161】

15 ECM関連タンパク質アッセイ
^{1 2 5}I-標識されたECM関連タンパク質の存在下で候補リガンド分子を用いてECMRP結合活性をリガンド結合アッセイで決定する。ECM関連タンパク質を、^{1 2 5}Iボルトンハンター試薬で標識する(Bolton A. E. および W. M. Hunter (1973) Biochem. J. 133: 529-539)。候補リガンド分子をマルチウェルプレートの各ウェルに注入し、標識したECM関連タンパク質と共にインキュベートし、次に洗浄し、標識したECM関連タンパク質複合体が存在するウェルをアッセイする。ECM関連タンパク質濃度を変えて得たデータを用いて、候補分子とのECM関連タンパク質の数、親和性及び会合の値を計算する。

【0162】

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。

【0163】

【図面の簡単な説明】

【図1A】

図はcDNA(SEQ ID NO: 2)によってコードされたECM関連タンパク質(SEQ ID NO: 1)を示す。アラインメントはMACDNASIS PROソフトウ

エア（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】

図はcDNA（SEQ ID NO：2）によってコードされたECM関連タンパク質（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1C】

図はcDNA（SEQ ID NO：2）によってコードされたECM関連タンパク質（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1D】

図はcDNA（SEQ ID NO：2）によってコードされたECM関連タンパク質（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】

図はcDNA（SEQ ID NO：2）によってコードされたECM関連タンパク質（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。

20

【図1F】

図はcDNA（SEQ ID NO：2）によってコードされたECM関連タンパク質（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1G】

図はcDNA（SEQ ID NO：2）によってコードされたECM関連タンパク質（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。

30

【図1H】

図はcDNA（SEQ ID NO：2）によってコードされたECM関連タンパク質（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1I】

図はcDNA（SEQ ID NO：2）によってコードされたECM関連タンパク質（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。

40

【図1J】

図はcDNA（SEQ ID NO：2）によってコードされたECM関連タンパク質（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1K】

図はcDNA（SEQ ID NO：2）によってコードされたECM関連タンパク質（SEQ ID NO：1）を示す。アラインメントはMACDNASIS PROソフトウ

50

エア（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2】

図2は多様な正常な成熟した組織のECMRPの発現を示す。X軸は組織のタイプ、Y軸は正常な結腸組織に見られるものに比較してECMRPの発現を示す（100%など）。分析はTAQMAN（Applied Biosystems，Foster City CA）によって実施された。

【図3】

図3は正常な結腸組織と比較して結腸癌を有する患者の組織のECMRPの示差発現を示す。X軸は患者ID（供与者ID）、Y軸は供与者ID3582の腫瘍組織での観察との比較で発現ECMRPを示す（例えば100%）。腫瘍サンプルは黒で表示され、正常な組織は白で表示される。分析はTAQMAN（Applied Biosystems）によって実施された。

10

【0166】

（表の簡単な説明）

表1は、LIFESEQ Gold database（Incyte Genomics，Palo Alto CA）を使用して作製したECMRP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はカテゴリのライブラリの総数に対して少なくとも1つの転写が見られるライブラリの数、列4は転写の絶対存在度（転写の数）、列5は転写の%存在度を示す。

20

【0167】

表2は、アレイ分析によって決定した正常な結腸と肺の組織に対する結腸と肺の癌を有する患者の組織のECMRPの示差発現を示す。列1は腫瘍サンプルと正常な組織の間の示差発現（DE）を示す。結果は腫瘍/正常発現の比という形で発現する。列2（P1説明）は、微視的には蛍光性緑色染料Cy3で標識する正常なサンプルの組織と患者供与者（Dn）を示す。列3（P2説明）は、蛍光性赤色染料Cy5で標識する病変サンプル（結腸腫瘍または結腸ポリープ、肺腫瘍）の組織と患者供与者（Dn）を示す。

【表1】

表1

組織	#cDNA	観察	存在度	%存在度
心血管系	270162	1/72	1	0.0004
結合組織	147886	1/54	1	0.0007
消化器系	514430	6/151	7	0.0014
胎芽構造	107325	2/23	2	0.0019
内分泌系	233587	3/63	4	0.0017
外分泌腺	255105	4/64	5	0.0020
女性生殖器	445078	3/113	3	0.0007
男性生殖器	453150	5/118	7	0.0015
生殖細胞	46185	0/5	0	0.0000
血液及び免疫系	701709	1/166	1	0.0001
肝臓	110945	1/34	1	0.0009
筋骨格系	162794	0/50	0	0.0000
神経系	973795	0/221	0	0.0000
脾臓	111757	2/25	3	0.0027
呼吸器系	407942	7/95	9	0.0022
感覚器	25346	0/10	0	0.0000
皮膚	72110	1/15	4	0.0055
顎口腔系	14025	1/11	1	0.0071
非分類性/混合性	150146	2/19	2	0.0013
尿路	287931	2/66	2	0.0007
総計	5321883	33/1292	55	0.0000

10

20

【表 2】

表2

DE	P1説明	P2説明
1.7	ヒト、結腸、Nrml, mw/アデノCA, Dn4614	ヒト、結腸腫瘍、アデノCA, Dn4614
2.3	ヒト、結腸ブール、Nrml, Dn3753	ヒト、結腸、ボリーブ、Dn3755
4.4	ヒト、結腸ブール、Nrml, Dn3753	ヒト、結腸腫瘍、癌、Dn3311
1.7	ヒト、結腸ブール、Nrml, Dn3753	ヒト、結腸、ボリーブ、Dn3754
2.5	ヒト、結腸ブール、Nrml, mw/アデノマ、Dn3583	ヒト、結腸腫瘍、アデノマ、Dn3583
3.3	ヒト、結腸、Nrml, mw/アデノCA, Dn3839	ヒト、結腸腫瘍、アデノCA, Dn3839
3.8	ヒト、結腸、直腸、Nrml, mw/癌、Dn3581	ヒト、結腸腫瘍、直腸、癌、Dn3581
1.5	ヒト、肺、Nrml, mw/扁平上皮細胞CA, Dn5796	ヒト、肺腫瘍、扁平上皮細胞CA, Dn5796
3.4	ヒト、肺、Nrml, mw/扁平上皮細胞CA, Dn5800	ヒト、肺腫瘍、扁平上皮細胞CA, Dn5800

10

20

30

【 1 A 】

5' ATG CTA TTG AGA GTC TTC CTC CAA ACA TCT TCC GAT TTG TTC CTT TAA CCC ATC 55
 10 19 28 37 46
 TAG ATC TTC GTG GAA ATC AAT TAC AAA CAT TGC CTT ATG TTG GTT TTC TCG AAC 109
 64 73 82 91
 ACA TTG GCC GAA TAT TGG ATC TTC AGT TGG AGG ACA ACA AAT GGG CCT GCA ATT 163
 118 127 136 145 154
 GTG ACT TAT TGC AGT TAA AAA CTT GGT TGG AGA ACA TGC CTC CAC AGT CTA TAA 217
 172 181 190 199 208
 CTT GGT GAT GTT CTC TGC AAC AGC CCT CCA TTT TTT AAA GGA AGT ATA CTC ZTT 211
 226 235 244 253 262
 AGA CTA AAG AAG GAA TCT ATT TGC CCT ACT CCA CCA GTG TAT GAA GAA CAT GAG 325
 280 289 298 307 316
 GAT CCT TCA GGA TCA TTA CAT CTG GCA GCA ACA TCT TCA ATA AAT GAT AGT CGC 379
 334 343 352 361 370

図 1A

【 1 B 】

ATG TCA ACT AAG ACC AGC TTC ATT CTA AAA CTA CCC ACC AAA GCA CCA GGT TTG 433
 388 397 406 415 424
 M S T K T T S I L K L P T K A P G L
 ATA CCT TAT AAT ACA AAG CCA TCC ACT GAA CTT CCA GGA CCT TAC TGC CTT AAT 487
 442 451 460 469 478
 I P Y I T K P S T Q L P G P Y C P I
 CCT TGT AAC TGC AAA GTC CTA TCC CCA TCA GGA CTT CTA ATA CAT TGT CAG GAG 541
 496 505 514 523 532
 P C N C K V L S P S G L L I H C Q E
 CGC AAC AAT GAA AGC TTA TCA GAT CTG AGA CCT CCT CCG GAA AAT CCT AGA AAG 595
 550 559 568 577 586
 R N I E S L S D L R P P P Q N P R K
 CTC AAT CTA GCG GGA AAT AAT AAT CAC AGT TTA ATG AAG TCT GAT CTA GTG GAA 649
 604 613 622 631 640
 L I L A G N I I H S L M K S D L L V E

図 1B

【 1 C 】

TAT TTC ACT TTG GAA ATG CTT CAC TTG GGA AAC AAT CGT APT GAA GPT CTT GAA 703
 667 676 685 694
 Y F T L E M L H L G N N R I E V L E
 GAA GGA TGS TTT ATG AAC CTA ACG AGA TTA CAA AAA CTC TAT CTA AAT GGT AAC 757
 712 721 730 739 748
 E G S F M N L T R L Q K L Y L N G N
 CAC CTG ACC AAA TTA AGT AAA GGC ATG TTC CTT GGT CTC CAT AAT CTT GAA TAC 811
 775 784 793 802 811
 H L T K L S K G M F L G L H N L E Y
 820 829 838 847 856 865
 TTA TAT CTT GAA TAC AAT GCC APT AAG GAA ATA CTG CCA GGA ACC TTT AAT CCA
 L Y L E Y N A I K E I S P G T F N P
 874 883 892 901 910 919
 ATG CCT AAA CTT AAA GTC CTG TAT TTA AAT AAC AAC CTC CTC CAA GTT TTA CCA
 M P K L K V L Y L N N N L L Q V L P

図 1C

【 1 D 】

CCA CAT AAT TTT TCA GGG GTT CCT CTA ACT AAG GTA AAT CTT AAA ACA AAC CAG 973
 937 946 955 964
 P H I F S G V P L T K V N L K T N Q
 TTT ACC CAT CTA CTT CTA AGT AAT AAT TTG GAT GAT CTT GAT TTA CTA ACC CAG 1027
 982 991 1000 1009 1018
 F T H L P V S N I L D D L L L L T Q
 AAT GAC CTT GAG GAT AAC CCC TGG GAC TGC TCC TGT GAC CTG GPT GGA CTG CAG 1081
 1036 1045 1054 1063 1072 1081
 I D L E D N P W D C S C D L V G L Q
 CAA TGG ATA CAA AAG TTA AGC AAG AAC ACA CTG ACA GAT GAC ATC CTC TGC ACT 1135
 1099 1108 1117
 Q W I Q K L S K N T V T D I L C Y
 1144 1153 1162 1171 1180 1189
 TCC CCC GGG CAT CTC GAC AAA AAG GAA TTG AAA GCC CTA AAT AGT GAA AAT CTC
 S P G H L D K K E L K A L N S E I L

図 1D

【 1 E 】

1198 1207 1216 1225 1234 1243
 TGT CCA GGT TTA GTA AAT AAC CCA TCC ATG CCA ACA CAG ACT AGT TAC CTT ATG
 C P G L V N N P S M P T Q T S Y L M

1252 1261 1270 1279 1288 1297
 CTC ACC ACT CCT GCA ACA ACA AAT AAT AGC GCT GAT ACT ATT TTA CGA TCT CTT
 V T T P A T T N T A D T I L R S L

1306 1315 1324 1333 1342 1351
 ACG GAC GCT GNG CCA CTG TCT GTT CTA ATA TTG GGA CTT CTG ATT ATG TTG ATC
 T D A V P L S V L I L G L L I M F I

1360 1369 1378 1387 1396 1405
 ACT ATT GTT TTC TGT GCT GCA GGG TAT CTG CTT CTT CTT CTT CTT CCG AGG AGA
 T I V F C A A G I V V L V L H R R R

1414 1423 1432 1441 1450 1459
 AGA TAC AAA AAG AAA CAA GTA GTA GAT GAG CAA ATG AGA GAC MAC AGT CCT GTC CRT
 R Y K K K Q V D E Q M R D N S P V H

1E

【 1 F 】

1468 1477 1486 1495 1504 1513
 CTT CAG TAC AGC ATG TAT GGC CAT AAA ACC ACT CAT CAC ACT ACT GAA AGA CCC
 L Q Y S M Y G H K T H T E R F

1522 1531 1540 1549 1558 1567
 TCT GCC TCA CTC TAT GAA CAG CAC ATG CTC AGC CCC ATG GTT CAT GTC TAT AGA
 S A S L Y E Q H M V S P M V H Y R

1576 1585 1594 1603 1612 1621
 AGT CCA TCC TTT GGT CCA AAG CAT CTG GAA GAG GAA GAA GAG AGG AAT GAG AAA
 S P S F G P K H L E E E E R N E K

1630 1639 1648 1657 1666 1675
 GAA GGA AGT GAT GCA AAA CAT CTC CAA AGA AGT CTT TTG GAA CAG GAA AAT CAT
 E G S D A K H L Q R S L E Q E N

1684 1693 1702 1711 1720 1729
 TCA CCA CTC ACA GGG TCA AAT ATG AAA TAC AAA ACC ACG AAC CAA TCA ACA GAA
 S P L T G S N M K Y K T T N Q S T E

1F

【 1 G 】

1738 1747 1756 1765 1774 1783
 TTT TTA TCC TTC CAA GAT GCC AGC TCA TTG TAC AGA AAC ATT TTA GAA AAA GAA
 F L S F Q D A S S L Y R N I L E K E

1792 1801 1810 1819 1828 1837
 AGS GAA CTT OAG CAA CTG GGA ATC ACA TAA TAC TAC AGG AAA AAC ATT GCT CAG
 R E L Q Q L G I T E Y L R K N I A Q

1846 1855 1864 1873 1882 1891
 CTC CAG CCT GAT ATC CAG GCA CAT TAT CCT GGA GCC CAC GAA GAG CTG AAG TTA
 L Q P D M E A H Y P G A H E E L K L

1900 1909 1918 1927 1936 1945
 ATG GAA CAA TTA AAG TTA TCA CTT CCA AGG AAG CTA TTA CTG TAA CAG ACA AAA
 M E T L M Y S R P R K V L V E Q T K

1954 1963 1972 1981 1990 1999
 AAT GAG TAT TTT GAA CTT AAA GCT AAT TTA CAT GCT GAA CCT GAC TAT TTA GAA
 N E Y F E L K A N L H A E F D Y L E

1G

【 1 H 】

2008 2017 2026 2035 2044 2053
 GTC CTG GAG CAG CAA ACA TAG ATG GAG AGT TTG AGG GCT TTC GCA GAA ATG CTG
 V L E Q T

2062 2071 2080 2089 2098 2107
 TGA TTC TGT TTT AAG TCC AFA CCT TGT AAA TAA GTG CCT TAC GTG AGT GTG TCA
 2116 2125 2134 2143 2152 2161
 TCA ATC AGA ACC TAA GCA CAG CAG TAA ACT ATG GCG AAA AAA GAA GAA GAA

2170 2179 2188 2197 2206 2215
 AAA GAA CTT CAG GGA TCA CTG GGA GAA GCC ATG GCA TTA TCT TCA GCC AAT TTA

2224 2233 2242 2251 2260 2269
 GTC TCT CCC AAA TAA AAT AAA TCC TTG CAT GTA AAT CMT TCA ACC GTT ATA GTA

2278 2287 2296 2305 2314 2323
 ATA TTT CAT ATA CTG AAA AGT GTC TCA TAG GAG TCC TCT TGC ACA TCT AAA AAG

1H

【 1 I 】

2332 GCT GAA CAT TTA AGT ATC CCG AAT TTT CTT GAA TTG CTT TCC CTA TAG ATT AAT 2368 2377
 2386 TAC AAT TGG AAT TCA TCA TTT AAA AAC CAT ACT TGT AFA TGT AGT TAT AAT ATG 2422 2431
 2440 TAA GCA ATA CAT TGT TTA TAA CCA GTA TGT ACT TCA AAA AUG TGT ATT CTC AAA 2476 2485
 2494 CAG ACC TAA CTT TCT TGC AAT AAA TCC AAA AGA AAC TGG AAC TTG ACA ATT ATA 2530 2539
 2548 AAT AGT AAT AGT GAA AAA ATA GAA AGG TTG GAA TTA TAA ACC GCA TGG TGG 2584 2593
 2602 GCT CAA AAC TTT GAA CAT TTG ACC TTA AAC AAA TGC CAC TCT CAT GCA TTC TAA 2638 2647

図 11

【 1 J 】

2656 AAT AAA AGC TTA AAA TGA TTA ATA GTT CAG GTG GAA GAA ATA AGC AFA CTT TTT 2701
 2710 GGG TTT TCT ACA CAT TTT GTG TAG ACA ATT TTA ATG TCA GTG CTG CTG TGA ACT 2746 2755
 2764 AAA GTA TGT CAT TTA TGC TCA AAG TTT AAT TCT TCT TCT TGG GAT ATT TTA AAA 2800 2809
 2818 ATG CTA CTG AGA TTC TGC TGT AAA TAT GAC TAG AGA ATA TAT TGG GTT TGC TTT 2854 2863
 2872 AAT TCA TGC GCT TAA TTC TTT GTA AAT CTG AAT GAC CAT AAT AGA AAT ACA TTT 2908 2917
 2926 CTT CTC GCA ACT AAT AAT TCA CAG TTG TAA AGT AAA TAG GAA AAA TTA TTT TAT TTT 2962 2971
 2980 TAT TCA TGC ACA TTC ATA GAT GCC ATA AAT CAG TAG CAA AAG GCA CTT CTA AAG 3016 3025

図 1J

【 1 K 】

3034 GTA AGT GGT TTA ACT TGC CTC AAG AGA GGG ACA ATG TAG CTT TAT TTT ACA AGA 3079
 3088 AGG CAT ACU TAG AAT TCT ATG AAA TAT TTA TTC TGT ACA GAT TTA TAT AAT TTT 3133
 3142 GGT TCA CAA AAG TAA TTA TTC TTG GGT GCC TTT CAA GAA AAT TAA AAA TAC TAC 3178 3187
 3196 CCA CTA CRA TAA AAC TAA AAT GAA AAC TCA AAA AAA A 3' 3223

図 1K

【 2 】

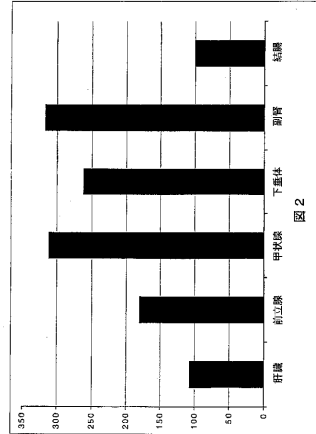
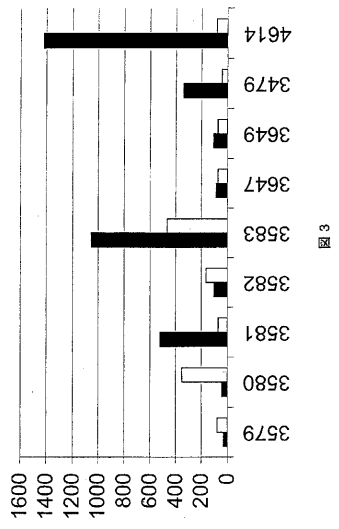


図 2

【 図 3 】



【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
10 January 2002 (10.01.2002)

PCT

(10) International Publication Number
WO 02/02772 A2

(51) International Patent Classification: C12N 15/12,
C07K 14/78, 16/18, C12Q 1/68, G01N 33/53, A61K
38/17, 39/395, A61P 35/00

W. [US/US]; 237 41st Street, #5, Oakland, CA 94611 (US).
YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA
94087 (US). PATTERSON, Chandra [US/US]; 490 Sher-
wood Way #1, Menlo Park, CA 94025 (US).

(21) International Application Number: PCT/US01/20729

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

(22) International Filing Date: 29 June 2001 (29.06.2001)

(25) Filing Language: English

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, 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,
TI, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(26) Publication Language: English

(30) Priority Data: 60/215,454 30 June 2000 (30.06.2000) US

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

(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, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BI, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, NI, SN, TD, TG).

(72) Inventors; and
(75) Inventors/Applicants (for US only): LASEK, Amy, K.,

[Continued on next page]

(54) Title: ECM-RELATED TUMOR MARKER



WO 02/02772 A2

```

10      19      28      37      46      55
5'  ATG CTA TTG AGA GTC TTC CTC CAA ACA TCT TCC GAT TTG TTC CTT TAA CCC ATC
-----
64      73      82      91      100     109
TAG ATC TFC GTG GAA AIC AAT TAC AAA CAT TGC CTT ATG TTG GAT TTC TGG AAC
-----
118     127     136     145     154     163
ACA ATG GTC GAA TAT TGG ATC TTC AGT TGG AGG ACA ACA AAT GGG CCF GCA APT
-----
172     181     190     199     208     217
GTG ACT TAT TGC AGT TAA AAA CTT GGT TGG AGA ACA TGC CTC CAC AGT CTA TAA
-----
226     235     244     253     262     271
CTT GGT GAT GAT GTC TGC AAC AGC CCT CCA TTT TTT AAA GGA AGT AFA CTC AGT
-----
280     289     298     307     316     325
AGA CTA AAG AAG GAA TCT ATT TGC CCT ACT CCA CCA GTG TAT GAA GAA CAT GAG
-----
334     343     352     361     370     379
GAT CCT TCA GGA TCA TTA CAT CTG GCA GCA ACA TCT TCA ATA AAT GAT AGT CGC
-----

```

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

WO 02/02772 A2



Published:

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

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

WO 02/02772

PCT/US01/20729

5

ECM-RELATED TUMOR MARKER

This application claims the benefit of provisional application U.S. Serial No. 60/215,454 filed June 30, 2000, all of which application is hereby incorporated by reference herein.

10

TECHNICAL FIELD

This invention relates to a cDNA which encodes an extracellular matrix(ECM)-related cancer marker and to the use of the cDNA and the encoded protein in the diagnosis and treatment of cancers, in particular, colon and lung cancer.

BACKGROUND OF THE INVENTION

15

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 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.

20

Cancers and malignant tumors are characterized by continuous cell proliferation and cell death and are causally related to both genetics and the environment. Cancer markers are of great importance in determining familial predisposition to cancers and in the early diagnosis and prognosis of various cancers.

25

Colorectal cancer is the fourth most common cancer and the second most common cause of cancer death in the United States with approximately 130,000 new cases and 55,000 deaths per year. Colon and rectal cancers share many environmental risk factors and both are found in individuals with specific genetic syndromes. (See Potter, JD (1999) J Natl Cancer Institute 91:916-932 for a review of colorectal cancer.) Colon cancer is the only cancer that occurs with approximately equal frequency in men and women, and the five-year survival rate following diagnosis of colon cancer is around 55% in the United States (Ries et al. (1990) National Institutes of Health, DHHS Publ No. (NIH)90-2789).

30

35

Colon cancer is causally related to both genes and the environment. Several molecular pathways have been linked to the development of colon cancer, and the expression of key genes in any of these pathways may be lost by inherited or acquired mutation or by hypermethylation. There is a particular need to identify genes for which changes in expression may provide an early indicator of colon cancer or a predisposition for the development of colon cancer, as well as potential

WO 02/02772

PCT/US01/20729

5 therapeutic targets for treatment of the disease.

A number of genes associated with the predisposition, development, and progression of colon cancer have been identified. For example, it is well known that abnormal patterns of DNA methylation occur consistently in human tumors. In colon cancer in particular, it has been found that these changes occur early in tumor progression such as in premalignant polyps that precede colon cancer. DNA methyltransferase, the enzyme that performs DNA methylation, is significantly increased in histologically normal mucosa from patients with colon cancer or the benign polyps that precede cancer, and this increase continues during the progression of colonic neoplasms (Wafik et al. (1991) Proc Natl Acad Sci USA 88:3470-3474). Familial Adenomatous Polyposis (FAP) is a rare autosomal dominant syndrome that precedes colon cancer and is caused by an inherited mutation in the adenomatous polyposis coli (APC) gene. The APC gene is a part of the APC- β -catenin-Tcf (T-cell factor) pathway. Impairment of this pathway results in the loss of orderly replication, adhesion, and migration of colonic epithelial cells that results in the growth of polyps. Hereditary nonpolyposis Colorectal Cancer (HNPCC) is another inherited autosomal dominant syndrome that is distinguished by the tendency to early onset of colon cancer and the development of other cancers. HNPCC results from the mutation of one or more genes in the DNA mis-match repair (MMR) pathway. Mutations in two human MMR genes, MSH2 and MLH1, are found in a large majority of HNPCC families identified to date. Almost all colon cancers arise from cells in which the estrogen receptor (ER) gene has been silenced. The silencing of ER gene transcription is age related and linked to hypermethylation of the ER gene (Issa et al. (1994) Nature Genetics 7:536-540). Introduction of an exogenous ER gene into cultured colon carcinoma cells results in marked growth suppression.

Clearly there are a number of genetic alterations associated with colon cancer and with the development and progression of the disease that potentially provide early indicators of cancer development, and which may also be used to monitor disease progression or provide therapeutic targets.

The extracellular matrix (ECM) is a complex network of glycoproteins, polysaccharides, proteoglycans, and other macromolecules that are secreted from a cell, and provides cells with a mechanical scaffold for adhesion, migration and signal transduction. Tumor cells often overcome the need for cell-ECM anchorage and ECM-mediated growth control in the process of becoming metastatic (Ruoslahti, E. (1996) Sci. Am. 275:72-77.) Thus alterations in ECM proteins provides an additional source of markers and potential therapeutic targets for some forms of cancer. Many ECM proteins are characterized by the presence of transmembrane domains and protein-protein or cell-matrix interaction domains such as leucine-rich repeats (LLR) believed to be associated with signal transduction and cellular adhesion as well as in protein-protein interactions (Gay, N.J., et al. (1991) FEBS Lett. 29:87-91).

WO 02/02772

PCT/US01/20729

5 The discovery of a cDNA encoding ECM-related protein satisfies a need in the art by providing compositions which are useful in the diagnosis and treatment of cancer particularly colon and lung cancer.

SUMMARY OF THE INVENTION

10 The invention is based on the discovery of a cDNA encoding ECM-related protein, ECMRP, which is useful in the diagnosis and treatment of cancer, particularly colon and lung cancer.

The invention provides an isolated cDNA comprising a nucleic acid sequence encoding a protein having the amino acid sequence of SEQ ID NO:1. The invention also provides an isolated cDNA or the complement thereof selected from the group consisting of a nucleic acid sequence of SEQ ID NO:2, a fragment of SEQ ID NO:2 selected from SEQ ID NOs:3-13, and a variant of SEQ ID NO:2 selected from SEQ ID NOs:14-15. The invention additionally provides a composition, a substrate, and a probe comprising the cDNA, or the complement of the cDNA, encoding ECM-related protein. The invention further provides a vector containing the cDNA, a host cell containing the vector and a method for using the cDNA to make ECM-related protein. The invention still further provides a transgenic cell line or organism comprising the vector containing the cDNA encoding ECM-related protein. The invention additionally provides a fragment, a variant, or the complement of the cDNA selected from the group consisting of SEQ ID NOs:2-15. In one aspect, the invention provides a substrate containing at least one of these fragments or variants or the complements thereof. In a second aspect, the invention provides a probe comprising a cDNA or the complement thereof which can be used in methods of detection, screening, and purification. In a further aspect, the probe is a single-stranded complementary RNA or DNA molecule.

25 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 cancer, in particular, colon and lung cancer. In another aspect, the cDNA or a fragment or a variant or the complements thereof may comprise an element on an array.

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

WO 02/02772

PCT/US01/20729

5 artificial chromosome constructions, peptides, transcription factors, repressors, and regulatory molecules.

The invention provides a purified protein or a portion thereof selected from the group consisting of an amino acid sequence of SEQ ID NO:1, a variant having at least 85% identity to the amino acid sequence of SEQ ID NO:1, an antigenic epitope of SEQ ID NO:1, and a biologically
10 active portion of SEQ ID NO:1. The invention also provides a composition comprising the purified protein and a pharmaceutical carrier. 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
15 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 cancer, in particular, colon and lung cancer.

The invention provides a method of using a protein to screen a subject sample for antibodies
20 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 a cancer, in particular, colon and lung cancer.

The invention also provides a method of using a protein to prepare and purify antibodies
25 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
30 expressed in a cancer, in particular, colon and lung cancer. The invention also provides a method of using an antibody to diagnose a cancer, in particular, colon and lung cancer comprising combining the antibody comparing the quantity of bound antibody to known standards, thereby establishing the presence of a cancer, in particular, colon and lung cancer. The invention further provides a method of
35 using an antibody to treat a cancer, in particular, colon and lung cancer comprising administering to a patient in need of such treatment a composition comprising the purified antibody and a pharmaceutical carrier.

The invention provides a method for inserting a heterologous marker gene into the genomic DNA of a mammal to disrupt the expression of the endogenous polynucleotide. The invention also

WO 02/02772

PCT/US01/20729

5 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:2-15, transforming the vector into an embryonic stem cell, selecting a transformed embryonic stem cell, 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
10 chimeric offspring containing the cDNA in its germ line, and breeding the chimeric mammal to produce a homozygous, mammalian model system.

BRIEF DESCRIPTION OF THE FIGURES AND TABLE

15 Figures 1A through 1K show the ECM-related protein (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).

Figure 2 shows the expression of ECMRP in various normal adult tissues. The X-axis indicates the tissue type, and the Y-axis the expression of ECMRP relative to that found in normal colon tissue (e.g., 100%). The analysis was performed by TAQMAN (Applied Biosystems, Foster City CA).

20 Figure 3 shows the differential expression of ECMRP in tissues from patients with colon cancer relative to normal colon tissue. The X-axis indicates the patient ID (Donor ID), and the Y-axis the expression ECMRP relative to that observed in tumor tissue from Donor ID 3582 (e.g., 100%). Tumor samples are displayed in black, and normal tissue in white. The analysis was performed by TAQMAN (Applied Biosystems).

25 Table 1 shows the Northern analysis for ECMRP produced using the LIFSEQ Gold 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 relative to the total number of libraries in that category; the fourth column, the absolute abundance of the transcript (number of transcripts); and the fifth column,
30 percent abundance of the transcript.

Table 2 shows the differential expression of ECMRP in tissues from patients with colon and lung cancer relative to normal colon or lung tissue as determined by array analysis. The first column lists the differential expression (DE) between the tumor sample and normal tissue. The results are expressed in terms of the ratio of tumor/normal expression. Column 2 (P1 Description) lists the
35 tissue and patient donor (Dn) for microscopically normal samples labeled with fluorescent green dye Cy3. Column 3 (P2 Description) lists the tissue and patient donor (Dn) for diseased samples (colon tumor or colon polyps, lung tumor) labeled with fluorescent red dye Cy5.

DESCRIPTION OF THE INVENTION

WO 02/02772

PCT/US01/20729

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

10 Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

Definitions

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

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

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

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

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

A "composition" refers to the polynucleotide and a labeling moiety, a purified protein and a

WO 02/02772

PCT/US01/20729

5 pharmaceutical carrier, an antibody and a labeling moiety, and the like.

"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, 10 formyl, or morpholino group. Derivative molecules retain the biological activities of the naturally occurring molecules but may confer advantages such as longer lifespan or enhanced activity.

"Differential expression" refers to an increased or upregulated or a decreased or downregulated expression as detected by absence, presence, or at least two-fold change in the amount of transcribed messenger RNA or translated protein in a sample.

15 "Disorder" refers to conditions, diseases or syndromes in which the cDNAs and ECM-related protein are differentially expressed. Such a disorder includes cancer, and in particular, colon and lung cancer.

"Fragment" refers to a chain of consecutive nucleotides from about 50 to about 4000 base pairs in length. Fragments may be used in PCR or hybridization technologies to identify related 20 nucleic acid molecules and in binding assays to screen for a ligand. Such ligands are useful as therapeutics to regulate replication, transcription or translation.

A "hybridization complex" is formed between a cDNA and a nucleic acid of a sample when the purines of one molecule hydrogen bond with the pyrimidines of the complementary molecule, e.g., 5'-A-G-T-C-3' base pairs with 3'-T-C-A-G-5'. Hybridization conditions, degree of 25 complementarity and the use of nucleotide analogs affect the efficiency and stringency of hybridization reactions.

"Labeling moiety" refers to any visible or radioactive label than can be attached to or incorporated into a cDNA or protein. Visible labels include but are not limited to anthocyanins, green fluorescent protein (GFP), β glucuronidase, luciferase, Cy3 and Cy5, and the like. Radioactive 30 markers include radioactive forms of hydrogen, iodine, phosphorous, sulfur, and the like.

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

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

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.

WO 02/02772

PCT/US01/20729

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

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

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

15 "Protein" refers to a polypeptide or any portion thereof. A "portion" of a protein refers to that length of amino acid sequence which would retain at least one biological activity, a domain identified by PFAM or PRINTS analysis or an antigenic epitope of the protein identified using Kyte-Doolittle algorithms of the PROTEAN program (DNASTAR, Madison WI).

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

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

"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
30 BLAST2 (Altschul *et al.* (1997) *Nucleic Acids Res* 25:3389-3402). BLAST2 may be used in a standardized and reproducible way to insert gaps in one of the sequences in order to optimize alignment and to achieve a more meaningful comparison between them. Particularly in proteins, similarity is greater than identity in that conservative substitutions, for example, valine for leucine or isoleucine, are counted in calculating the reported percentage. Substitutions which are considered to
35 be conservative are well known in the art.

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

WO 02/02772

PCT/US01/20729

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

"Variant" refers to molecules that are recognized variations of a cDNA or a protein encoded by the cDNA. Splice variants may be determined by BLAST score, wherein the score is at least 100, and most preferably at least 400. Allelic variants have a high percent identity to the cDNAs and may differ by about three bases per hundred bases. "Single nucleotide polymorphism" (SNP) refers to a 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 ECM-related protein and on the use of the cDNA, or fragments thereof, and protein, or portions thereof, directly or as compositions in the characterization, diagnosis, and treatment of colon and lung cancer.

20 Nucleic acids encoding the ECM-related protein of the present invention were first identified in Incyte Clone 2743093 using a computer search for nucleotide and/or amino acid sequence alignments. SEQ ID NO:2 was derived from the following overlapping and/or extended nucleic acid sequences (SEQ ID NO:3-13): Incyte Clones 2743093H1 (SKINDIA01), 4876623F9 (COLDNOT01), 2316239T6 (OVARNOT02), 6258015F8 (BMARTXT06), and shotgun sequences

25 7677606J1, 71111915V1, 71112850V1, 71262960V1, 71264035V1, 71113484V1, and 71114738V1. In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1 as shown in Figures 1A through 1J. ECM-related protein is 546 amino acids in length and has two potential N-glycosylation sites at N114 and N446; eight potential casein kinase II phosphorylation sites at S59, T215, T244, S305, S381, S425, S453, and T521; two potential protein kinase C phosphorylation sites at S2 and T375; and one potential tyrosine kinase phosphorylation site at Y510. PFAM analysis indicates several leucine rich repeat domains (LRR) associated with protein-protein binding interactions between amino acid residues T93 and P188 of SEQ ID NO:1. PFAM analysis further indicates a C-terminal LLR from residues N222 through P272. PRINTS analysis further confirms the presence of LLRs from L118 to L131 and from M163 to L176 of SEQ ID NO:1. HMMR analysis further indicates the presence of a transmembrane domain between residues I316 and V335 of SEQ ID NO:1. A useful antigenic epitope extends from about L60 to about L175 of ECMRP and biologically active portions of ECMRP extend from about L118 to about L131, and from about M163 to about L176. An antibody which specifically binds ECM-related protein is useful in a diagnostic assay to identify a cancer, in particular colon and lung cancer.

WO 02/02772

PCT/US01/20729

5 Figure 2 shows the results of various normal adult tissues analyzed for ECMRP expression by TAQMAN analysis. Significant expression of the ECMRP was found only in liver, prostate, colon, thyroid, pituitary and adrenal tissues, and was undetectable in heart, brain, lung, skeletal muscle, kidney, pancreas, spleen, thymus, ovary, small intestine, and peripheral blood leukocytes.

10 Table 1 shows the expression of the ECMRP across tissue categories by Northern analysis of cDNA libraries in the LifeSeq database (Incyte Genomics). The results show the expression of ECMRP distributed across a variety of tissue categories including digestive system, endocrine and exocrine system, male genitalia (e.g., prostate), respiratory system, and skin. The differences observed between the results of Table 1 and Figure 2, above, most likely reflect the high incidence of fetal and diseased tissues in cDNA libraries of the LifeSeq database.

15 Figure 3 shows the expression of ECMRP in colon cancer tissue samples compared with normal colon tissue and in lung tumor samples compared with normal lung performed by TAQMAN analysis (Applied Biosystems). The results show an increased expression of ECMRP in colon tumors in 6 out of 9 samples examined (Donor IDs:3581, 3583, 3647, 3649, 3479, and 4614). The results were considered significant if at least a 1.2-fold difference in expression between cancerous and normal tissue was observed.

20 Table 2 shows the results of microarray analysis comparing the expression of ECMRP in colon cancer or colon polyp tissues relative to normal colon tissue, and in lung tumor tissue relative to normal lung. The results show an increased expression of ECMRP in colon tumors or polyps in 7 of 14 patients examined (Donor IDs:4614, 3755, 3311, 3754, 3583, 3839, and 3581). Increased expression of ECMRP was also observed in 2 of 10 lung tumor samples examined (Donor IDs:5796 and 5800). Differential expression (column 2) was considered significant if at least a 1.5-fold difference in expression between cancerous and normal tissue was observed.

Increased expression of ECMRP was also observed in a human colorectal adenocarcinoma cell line, HT-29, when compared to normal colon tissue. (Data not shown)

30 Mammalian variants of the cDNA encoding ECM-related protein were identified using BLAST2 with default parameters and the ZOOSEQ databases (Incyte Genomics). These preferred variants have from about 87% to 89% identity as shown in the table below. The first column, the SEQ ID_{var} 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 _{var}	cDNA _{var}	Species	Identity	N _H Alignment
14	703545995J1	Dog	89%	2-192
15	112357_Mm.1	Monkey	87%	3003-3173

40

WO 02/02772

PCT/US01/20729

5 It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of cDNAs encoding ECM-related protein, 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 ECM-related protein, and all such variations are to be considered as being specifically disclosed.

10 The cDNAs of SEQ ID NOs:2-15 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, SEQ ID NOs:14-15, may be used to produce transgenic cell lines or organisms which are model systems for human colon or lung cancer and upon which the toxicity and efficacy of potential therapeutic treatments may be tested. Toxicology studies, clinical trials, and subject/patient treatment profiles may be performed and monitored using the cDNAs, proteins, antibodies and molecules and compounds identified using the cDNAs and proteins of the present invention.

15 The identification and characterization of the cDNAs and proteins, fragments or portions thereof, were described in provisional application U.S. Serial No. 60/215,454, incorporated by reference herein in its entirety.

Characterization and Use of the Invention

cDNA Libraries

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

Sequencing

25 Methods for sequencing nucleic acids are well known in the art and may be used to practice any of the embodiments of the invention. These methods employ enzymes such as the Klenow fragment of DNA polymerase I, SEQUENASE, Taq DNA polymerase and thermostable T7 DNA polymerase (Amersham Pharmacia Biotech (APB), Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such

WO 02/02772

PCT/US01/20729

5 as the MICROLAB 2200 system (Hamilton, Reno NV) and the DNA ENGINE thermal cycler (MJ Research, Watertown MA). Machines commonly used for sequencing include the ABI PRISM 3700, 377 or 373 DNA sequencing systems (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (APB), and the like. The sequences may be analyzed using a variety of algorithms well known 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).

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

20 Contaminating sequences, including vector or chimeric sequences, or deleted sequences can be removed or restored, respectively, organizing the incomplete assembled sequences into finished sequences.

Extension of a Nucleic Acid Sequence

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

Hybridization

The cDNA and fragments thereof can be used in hybridization technologies for various purposes. A probe may be designed or derived from unique regions such as the 5' regulatory region or from a nonconserved region (i.e., 5' or 3' of the nucleotides encoding the conserved catalytic domain of the protein) and used in protocols to identify naturally occurring molecules encoding the ECM-related protein, 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,

WO 02/02772

PCT/US01/20729

5 SEQ ID NOs:2-15. Hybridization probes may be produced using oligolabeling, nick translation, end-labeling, or PCR amplification in the presence of a reporter molecule. A vector containing the cDNA or a fragment thereof may be used to produce an mRNA probe in vitro by addition of an RNA polymerase and labeled nucleotides. These procedures may be conducted using commercially available kits such as those provided by APB.

10 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. 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
15 mismatches. Subsequent washes are performed at higher stringency with buffers such as 0.2xSSC with 0.1% SDS at either 45C (medium stringency) or 68C (high stringency). At high stringency, hybridization complexes will remain stable only where the nucleic acids are completely complementary. In some membrane-based hybridizations, preferably 35% or most preferably 50%, formamide can be added to the hybridization solution to reduce the temperature at which
20 hybridization is performed, and background signals can be reduced by the use of detergents such as Sarkosyl or TRITON X-100 (Sigma-Aldrich, St. Louis MO) and a blocking agent such as denatured salmon sperm DNA. Selection of components and conditions for hybridization are well known to those skilled in the art and are reviewed in Ausubel (supra) and Sambrook et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY.

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

Hybridization probes are also useful in mapping the naturally occurring genomic sequence. The probes may be hybridized to a particular chromosome, a specific region of a chromosome, or an
35 artificial chromosome construction. Such constructions include human artificial chromosomes (HAC), yeast artificial chromosomes (YAC), bacterial artificial chromosomes (BAC), bacterial P1 constructions, or the cDNAs of libraries made from single chromosomes.

Expression

WO 02/02772

PCT/US01/20729

5 Any one of a multitude of cDNAs encoding ECM-related protein 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
10 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).

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

Routine cloning, subcloning, and propagation of nucleic acid sequences can be achieved
25 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.

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

WO 02/02772

PCT/US01/20729

5 hybridizations or PCR amplification techniques.

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

Recovery of Proteins from Cell Culture

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

WO 02/02772

PCT/US01/20729

5 thereof may be 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

10 Various hosts including, but not limited to, goats, rabbits, rats, mice, and human cell lines may be immunized by injection with ECM-related protein 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, 15 which are identical to a portion of the natural protein. Oligopeptides may be fused with proteins such as KLH in order to produce antibodies to the chimeric molecule.

Monoclonal antibodies may be prepared using any technique which provides for the production of antibodies by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. 20 (See, e.g., Kohler et al. (1975) *Nature* 256:495-497; Kozbor et al. (1985) *J. Immunol Methods* 81:31-42; Cote et al. (1983) *Proc Natl Acad Sci* 80:2026-2030; and Cole et al. (1984) *Mol Cell Biol* 62:109-120.)

Alternatively, techniques described for antibody production may be adapted, using methods known in the art, to produce epitope-specific, single chain antibodies. Antibody fragments which 25 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, e.g., Huse et al. (1989) 30 *Science* 246:1275-1281.)

The ECM-related protein 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 typically 35 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).

WO 02/02772

PCT/US01/20729

5 Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for ECMRP. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of HSPDE10A-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple epitopes, represents the average affinity, or avidity, of the antibodies for ECMRP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} l/mole are preferred for use in immunoassays in which the protein-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 l/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of ECMRP, preferably in active form, from the antibody (Catty (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell and Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley, New York NY). Labeling of Molecules for Assay

20 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 32 P-dCTP (APB), Cy3-dCTP or Cy5-dCTP (Operon Technologies, Alameda CA), or amino acid such as 35 S-methionine (APB). Nucleotides and amino acids may be directly labeled with a variety of substances including fluorescent, chemiluminescent, or chromogenic agents, and the like, by chemical conjugation to amines, thiols and other groups present in the molecules using reagents such as BIODIPY or FITC (Molecular Probes, Eugene OR).

DIAGNOSTICS

Nucleic Acid Assays

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

35 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.

WO 02/02772

PCT/US01/20729

5 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
10 expression profiles are established. This is accomplished by combining a sample taken from normal subjects, either animal or human, with a cDNA under conditions for hybridization to occur. Standard hybridization complexes may be quantified by comparing the values obtained using normal subjects with values from an experiment in which a known amount of a purified sequence is used. Standard values obtained in this manner may be compared with values obtained from samples from patients
15 who were diagnosed with a particular condition, disease, or disorder. Deviation from standard values toward those associated with a particular disorder is used to diagnose that disorder.

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

Protein Assays

Detection and quantification of a protein using either labeled amino acids or specific
25 polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include two-dimensional polyacrylamide gel electrophoresis, enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). These assays and their quantitation against purified, labeled standards are well known in the art (Ausubel, supra, unit 10.1-10.6). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two
30 non-interfering epitopes is preferred, but a competitive binding assay may be employed. (See, e.g., Coligan et al. (1997) Current Protocols in Immunology, Wiley-Interscience, New York NY; and Pound, supra.)

THERAPEUTICS

Chemical and structural similarity, in particular the a transmembrane domain and several
35 LLR domains, exists between regions of ECM-related protein (SEQ ID NO:1) and other ECM-related proteins. In addition, differential expression is highly associated with colon and lung cancer as shown in Table 2 and Figure 3. ECM-related protein clearly plays a role in cancer, in particular, colon and lung cancer.

WO 02/02772

PCT/US01/20729

5 In the treatment of conditions associated with increased expression of the protein such as colon or lung cancer, 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 and a pharmaceutical carrier may be administered to a subject to treat a condition associated with the increased expression or activity of the endogenous protein. In an additional embodiment, a vector expressing the complement of the cDNA or fragments thereof may be administered to a subject to treat the disorder.

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

Modification of Gene Expression Using Nucleic Acids

20 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 ECM-related protein. Oligonucleotides designed to inhibit transcription initiation are preferred. Similarly, inhibition can be achieved using triple helix base-pairing which inhibits the binding of polymerases, transcription factors, or regulatory molecules (Gee *et al.* In: Huber and Carr (1994) Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary molecule may also be designed to block translation by preventing binding between ribosomes and mRNA. In one alternative, a library or plurality of cDNAs may be screened to identify those which specifically bind a regulatory, nontranslated sequence.

30 Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA followed by endonucleolytic cleavage at sites such as GUA, GUU, and GUC. Once such sites are identified, an oligonucleotide with the same sequence may be evaluated for secondary structural features which would render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing their hybridization with complementary oligonucleotides using ribonuclease protection assays.

35 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

WO 02/02772

PCT/US01/20729

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

- 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.

- 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 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 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 a chaotropic agent to separate the protein from the purified ligand.

In a preferred embodiment, ECM-related protein may be used to screen a plurality of

WO 02/02772

PCT/US01/20729

5 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
10 plurality or libraries of ligands, and the specificity of binding or formation of complexes between the expressed protein and the ligand can be measured. Depending on the particular kind of molecules or compounds 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.

15 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
20 specifically compete with a test compound capable of binding to the protein. Molecules or compounds identified by screening may be used in a mammalian model system to evaluate their toxicity, diagnostic, or therapeutic potential.

Pharmacology

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

30 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., ED₅₀ (the dose therapeutically effective in 50% of the population) and LD₅₀ (the dose lethal to 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index, and it
35 may be expressed as the ratio, LD₅₀/ED₅₀. 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.

Model Systems

WO 02/02772

PCT/US01/20729

5 Animal models may be used as bioassays where they exhibit a phenotypic response similar to that of humans and where exposure conditions are relevant to human exposures. Mammals are the most common models, and most infectious agent, cancer, drug, and toxicity studies are performed on rodents such as rats or mice because of low cost, availability, lifespan, reproductive potential, and abundant reference literature. Inbred and outbred rodent strains provide a convenient model for
10 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.

Toxicology

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

20 Genetic toxicology identifies and analyzes the effect of an agent on the rate of endogenous, spontaneous, and induced genetic mutations. Genotoxic agents usually have common chemical or physical properties that facilitate interaction with nucleic acids and are most harmful when chromosomal aberrations are transmitted to progeny. Toxicological studies may identify agents that increase the frequency of structural or functional abnormalities in the tissues of the progeny if
25 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.

Acute toxicity tests are based on a single administration of an agent to the subject to determine the symptomology or lethality of the agent. Three experiments are conducted: 1) an initial
30 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
35 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

WO 02/02772

PCT/US01/20729

5 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 10 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

15 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 20 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 25 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.

30 Knockout Analysis

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 35 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.

WO 02/02772

PCT/US01/20729

5 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
10 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
15 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
20 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
25 "extensive metabolizers" to "poor metabolizers" of these agents.

In additional embodiments, the cDNAs which encode the 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.

30

EXAMPLES

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention. The preparation of the human colon polyp library, COLDNOT01, is described.

35 **I cDNA Library Construction**COLDNOT01

The COLDNOT01 library was constructed using RNA isolated from diseased descending colon tissues from a 16-year-old male during partial colectomy, temporary ileostomy, and

WO 02/02772

PCT/US01/20729

5 colonoscopy. Pathology indicated innumerable (greater than 100) adenomatous polyps with low grade dysplasia involving the entire colonic mucosa in the setting of familial polyposis coli.

The frozen tissue was homogenized and lysed in guanidinium isothiocyanate solution using a POLYTRON homogenizer (Brinkmann Instruments, Westbury NJ). The lysate was centrifuged over a 5.7 M CsCl cushion using an SW28 rotor in an L8-70M ultracentrifuge (Beckman Coulter, Fullerton CA) for 18 hours at 25,000 rpm at ambient temperature. The RNA was extracted with acid phenol, pH 4.7, precipitated using 0.3 M sodium acetate and 2.5 volumes of ethanol, resuspended in RNAse-free water, and DNase treated at 37°C. Extraction with acid phenol, pH 4.7, and precipitation with sodium acetate and ethanol was repeated. The mRNA was isolated with the OLIGOTEX kit (Qiagen, Chatsworth CA) and used to construct the cDNA library.

15 The mRNA was handled according to the recommended protocols in the SUPERSERIPIT plasmid system (Life Technologies) which contains a NotI primer-adaptor designed to prime the first strand cDNA synthesis at the poly(A) tail of mRNAs. Double stranded cDNA was blunted, ligated to EcoRI adaptors and digested with NotI (New England Biolabs, Beverly MA). The cDNAs were fractionated on a SEPHAROSE CL4B column (APB), and those cDNAs exceeding 400 bp were
20 ligated into pINCY plasmid (Incyte Genomics). The plasmid pINCY was subsequently transformed into DH5 α competent cells (Life Technologies).

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
25 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.

An intermediate plasmid, pSPORT 1- Δ RI, which showed no digestion with EcoRI, was digested with Hind III (New England Biolabs); and the overhanging ends were filled in with Klenow and dNTPs. A linker sequence was phosphorylated, ligated onto the 5' blunt end, digested with
30 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 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
35 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 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

WO 02/02772

PCT/US01/20729

5 Biosystems, Gaithersburg MD) or the REAL PREP 96 plasmid kit (Qiagen). A kit consists of a 96-well block with reagents for 960 purifications. The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile 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
10 isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were transferred to a 96-well block for storage at 4C.

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
15 sequencing system (Applied Biosystems) or the MEGABACE 1000 DNA sequencing system (APB). Most of the isolates were sequenced according to standard ABI protocols and kits (Applied Biosystems) with solution volumes of 0.25x-1.0x concentrations. In the alternative, cDNAs were sequenced using solutions and dyes from APB.

IV Extension of cDNA Sequences

20 The cDNAs were extended using the cDNA clone and oligonucleotide primers. One primer was synthesized to initiate 5' extension of the known fragment, and the other, to initiate 3' extension of the known fragment. The initial primers were designed using commercially available primer analysis software to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68C to about 72C. Any stretch of
25 nucleotides that would result in hairpin structures and primer-primer dimerizations was avoided.

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

High fidelity amplification was obtained by PCR using methods such as that taught in USPN 5,932,451. PCR was performed in 96-well plates using the DNA ENGINE thermal cycler (MJ Research). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg²⁺, (NH₄)₂SO₄, and β-mercaptoethanol, Taq DNA polymerase (APB), ELONGASE
35 enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B (Incyte Genomics): Step 1: 94C, three min; Step 2: 94C, 15 sec; Step 3: 60C, one min; Step 4: 68C, two min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68C, five min; Step 7: storage at 4C. In the alternative, the parameters for primer pair T7 and SK+

WO 02/02772

PCT/US01/20729

5 (Stratagene) were as follows: Step 1: 94C, three min; Step 2: 94C, 15 sec; Step 3: 57C, one min; Step 4: 68C, two min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68C, five min; Step 7: storage at 4C.

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

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

20 The cells were lysed, and DNA was amplified using primers, Taq DNA polymerase (APB) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94C, three min; Step 2: 94C, 15 sec; Step 3: 60C, one min; Step 4: 72C, two min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72C, five min; Step 7: storage at 4C. DNA was quantified using PICOGREEN quantitation reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the conditions described above. Samples were diluted with 20% dimethylsulfoxide (DMSO; 30 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

35 The cDNAs of the Sequence Listing or their deduced amino acid sequences were used to query databases such as GenBank, SwissProt, BLOCKS, and the like. These databases that contain previously identified and annotated sequences or domains were searched using BLAST or BLAST2 to produce alignments and to determine which sequences were exact matches or homologs. The alignments were to sequences of prokaryotic (bacterial) or eukaryotic (animal, fungal, or plant)

WO 02/02772

PCT/US01/20729

5 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 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 and Altschul (1993; Proc Natl Acad Sci 90:5873-5877), BLAST
10 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
15 comparison with hybridization procedures used in the laboratory, the stringency for an exact match was set from a lower limit of about 40 (with 1-2% error due to uncalled bases) to a 100% match of about 70.

The BLAST software suite (NCBI, Bethesda MD; <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>), includes various sequence analysis programs including
20 "blastn" that is used to align nucleotide sequences and BLAST2 that is used for direct pairwise comparison of either nucleotide or amino acid sequences. BLAST programs are commonly used with gap and other parameters set to default settings, e.g.: Matrix: BLOSUM62; Reward for match: 1; Penalty for mismatch: -2; Open Gap: 5 and Extension Gap: 2 penalties; Gap x drop-off: 50; Expect: 10; Word Size: 11; and Filter: on. Identity is measured over the entire length of a sequence. Brenner
25 *et al.* (1998; Proc Natl Acad Sci 95:6073-6078, incorporated herein by reference) analyzed BLAST for its ability to identify structural homologs by sequence identity and found 30% identity is a reliable threshold for sequence alignments of at least 150 residues and 40%, for alignments of at least 70 residues.

The cDNAs of this application were compared with assembled consensus sequences or
30 templates found in the LIFESEQ GOLD database (Incyte Genomics). Component sequences from cDNA, extension, full length, and shotgun sequencing projects were subjected to PHRED analysis and assigned a quality score. All sequences with an acceptable quality score were subjected to various pre-processing and editing pathways to remove low quality 3' ends, vector and linker sequences, polyA tails, Alu repeats, mitochondrial and ribosomal sequences, and bacterial
35 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

WO 02/02772

PCT/US01/20729

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

Bins were compared to one another, and those having local similarity of at least 82% were combined and reassembled. Bins having templates with less than 95% local identity were split. Templates were subjected to analysis by STITCHER/EXON MAPPER algorithms that determine the
15 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
20 an E-value (or probability score) of $\leq 1 \times 10^{-8}$. The templates were also subjected to frameshift FASTx against GENPEPT, and homolog match was defined as having an E-value of $\leq 1 \times 10^{-8}$. Template analysis and assembly was described in USSN 09/276,534, filed March 25, 1999.

Following assembly, templates were subjected to BLAST, motif, and other functional analyses and categorized in protein hierarchies using methods described in USSN 08/812,290 and
25 USSN 08/811,758, both filed March 6, 1997; in USSN 08/947,845, filed October 9, 1997; and in USSN 09/034,807, filed March 4, 1998. Then templates were analyzed by translating each template in all three forward reading frames and searching each translation against the PFAM database of hidden Markov model-based protein families and domains using the HMMER software package (Washington University School of Medicine, St. Louis MO; <http://pfam.wustl.edu/>). The cDNA was
30 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

35 Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon 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 ECM-related protein that have been mapped

WO 02/02772

PCT/US01/20729

5 result in the assignment of all related regulatory and coding sequences to the same location. The genetic map locations are described as ranges, or intervals, of human chromosomes. The map position of an interval, in cM (which is roughly equivalent to 1 megabase of human DNA), is measured relative to the terminus of the chromosomal p-arm.

VII Hybridization Technologies and Analyses

10 Tissue Sample Preparation

Matched normal colon and cancerous colon or colon polyp tissue samples were provided by the Huntsman Cancer Institute, (Salt Lake City, UT), and are described as follows: Donor 3754, age and sex unknown, pendunculated colon polyp; Donor 3755, age and sex unknown, polyp, family history of colon cancer; Donor 3583, 58 years old, male, tubulovillous adenoma (polyp), patient was also diagnosed with a hyperplastic polyp; Donor 3311, 85 years old, male, invasive, poorly differentiated adenocarcinoma, metastatic, 2/9 lymph nodes positive, TNM classification: T4, N1, Mx, patient was also diagnosed with multiple tubular adenomas; Donor 3839, 60 years old, sex unknown, colon cancer, no pathology report; Donor 4614, 67 years old, sex unknown, colon adenocarcinoma, moderately differentiated, DUKE'S B, TNM classification: T3, N0. In Table 2, Donor samples 3754, 3755, and 3311 were compared against a common control tissue designated Donor 3753, which was a pool of normal colon tissue from 3 donors. All other comparisons were done with matched normal and tumor tissue from the same donor.

Matched normal and lung tumor tissue samples (Donor IDs 5796 and 5800) were obtained from the Roy Castle International Institute for Cancer Research (Liverpool, United Kingdom). Donor 5796 is a 66 year-old male with a squamous cell carcinoma. Donor 5800 is a 75 year-old female with a squamous cell carcinoma.

The following normalized, first-strand, cDNA preparations of human, adult, normal tissues were obtained from Clontech Laboratories Inc. (Palo Alto CA): heart, pooled from 16 male/female Caucasians, ages 25-59; ; brain (whole), pooled from 4 male Caucasians, ages 43-55; placenta, pooled from 10 female Caucasians, ages 22-35; lung, pooled from 2 female Caucasians, ages 24 and 32; liver, 35-yr-old male Caucasian; skeletal muscle, pooled from 35 male/female Caucasians, ages 20-60; kidney, pooled from 14 male/female Caucasians, ages 24-55; pancreas, pooled from 20 male/female Caucasians, ages 25-59; spleen, pooled from 6 male/female Caucasians, ages 24-61; thymus, pooled from 9 male/female Caucasians, ages 18-32; prostate, pooled from 20 Caucasians, ages 20-58; testis, pooled from 45 Caucasians, ages 14-64; ovary, pooled from 7 Caucasians, ages 17-60; small intestine, pooled from 32 male/female Caucasians, ages 15-57; colon, pooled from 20 male/female Caucasians, ages 17-76 (includes inner mucosal lining); and peripheral blood leukocytes, pooled from male/female Caucasians, ages 18-40, all samples negative for HIV-I, HIV-II,

WO 02/02772

PCT/US01/20729

5 hepatitis B and syphilis.

The HT-29 human colorectal adenocarcinoma cell line was obtained from The American Type Culture Collection (ATCC, Manassas VA) and was cultured according to the suppliers specifications.

Immobilization of cDNAs on a Substrate

10 The cDNAs are applied to a substrate by one of the following methods. A mixture of cDNAs is fractionated by gel electrophoresis and transferred to a nylon membrane by capillary transfer. Alternatively, the cDNAs are individually ligated to a vector and inserted into bacterial host cells to form a library. The cDNAs are then arranged on a substrate by one of the following methods. In the first method, bacterial cells containing individual clones are robotically picked and arranged on a
15 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 irradiated in a
20 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 μ g. Amplified nucleic acids from about 400 bp to about 5000 bp in length are purified using SEPHACRYL-400 beads
25 (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%
30 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
35 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 TAQMAN Analysis

WO 02/02772

PCT/US01/20729

5 Probes for TAQMAN analysis were prepared according to the manufacturer's protocol (Applied Biosystems).

Probe Preparation for Membrane Hybridization

Hybridization probes derived from the cDNAs of the Sequence Listing are employed for screening cDNAs, mRNAs, or genomic DNA in membrane-based hybridizations. Probes are prepared by diluting the cDNAs to a concentration of 40-50 ng in 45 μ l TE buffer, denaturing by heating to 100C 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 37C 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 100C 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

The following method was used for the preparation of probes for the microarray analysis presented in Table 2. 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, 1:1000, and 1:100 (w/w) to sample mRNA respectively. To examine mRNA differential expression patterns, a second set of control mRNAs are diluted into reverse transcription reaction mixture at ratios of 1:3, 3:1, 1:10, 10:1, 1:25, and 25:1 (w/w). The reaction mixture is mixed and incubated at 37C for two hr. The reaction mixture is then incubated for 20 min at 85C, and probes are purified using two successive CHROMA SPIN+TE 30 columns (Clontech, Palo Alto CA). Purified probe is ethanol precipitated by diluting probe to 90 μ l in DEPC-treated water, adding 2 μ l 1mg/ml glycogen, 60 μ l 5 M sodium acetate, and 300 μ l 100% ethanol. The probe is centrifuged for 20 min at 20,800xg, and the pellet is resuspended in 12 μ l resuspension buffer, heated to 65C for five min, and mixed thoroughly. The probe is heated and mixed as before and then stored on ice. Probe is used in high density array-based hybridizations as described below.

Membrane-based Hybridization

WO 02/02772

PCT/US01/20729

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

Polymer Coated Slide-based Hybridization

The following method was used in the microarray analysis presented in Table 2. Probe is heated to 65C for five min, centrifuged five min at 9400 rpm in a 5415C microcentrifuge (Eppendorf Scientific, Westbury NY), and then 18 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 60C. The arrays are washed for 10 min at 45C in 1xSSC, 0.1% SDS, and three times for 10 min each at 45C in 0.1xSSC, and dried.

Hybridization reactions are performed in absolute or differential hybridization formats. In the absolute hybridization format, probe from one sample is hybridized to array elements, and signals are detected after hybridization complexes form. Signal strength correlates with probe mRNA levels in 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 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. Filters positioned between the array and the photomultiplier

WO 02/02772

PCT/US01/20729

5 tubes are used to separate 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.

10 The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping
15 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
20 analysis is the GEMTOOLS program (Incyte Genomics).

Solution-based Hybridization (TAQMAN)

Hybridization reactions and detection of hybridization complexes for TAQMAN analysis were performed according to the manufacturer's protocol (Applied Biosystems).

VIII Electronic Analysis

25 BLAST was used to search for identical or related molecules in the GenBank or LIFEBSEQ 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
30 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 and is shown in Figure 3.

35 All sequences and cDNA libraries in the LIFEBSEQ 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,

WO 02/02772

PCT/US01/20729

5 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, significant expression may reflect presence or absence or differential expression of the cDNA.

IX Complementary Molecules

10 Molecules complementary to the cDNA, from about 5 (PNA) to about 5000 bp (complement of a cDNA insert), are used to detect or inhibit gene expression. Detection is described in Example 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
15 sequences (such as enhancers or introns) and are used in "triple helix" base pairing to compromise the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. To inhibit translation, a complementary molecule is designed to prevent ribosomal binding to the mRNA encoding the protein.

20 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 elements for inducing vector replication are used in the transformation/expression system.

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

30 X Expression of ECM-related protein

Expression and purification of the protein are achieved using either a mammalian cell expression system or an insect cell expression system. The pUB6/V5-His vector system (Invitrogen, Carlsbad CA) is used to express ECM-related protein in CHO cells. The vector contains the selectable bsd gene, multiple cloning sites, the promoter/enhancer sequence from the human ubiquitin
35 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

WO 02/02772

PCT/US01/20729

5 californica nuclear polyhedrosis virus (baculovirus). The polyhedrin gene is replaced with the 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

10 ECM-related protein 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 ECM-related protein 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
15 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-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
20 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.

25 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
30 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-
35 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

WO 02/02772

PCT/US01/20729

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

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

XV ECM-related protein Assay

- ECMRP binding activity is determined in a ligand-binding assay using candidate ligand
35 molecules in the presence of 125 I-labeled ECM-related protein. ECM-related protein is labeled with 125 I Bolton-Hunter reagent (Bolton and Hunter (1973) Biochem J 133:529-539). Candidate ligand molecules, previously arrayed in the wells of a multi-well plate, are incubated with the labeled ECM-related protein, washed, and any wells with labeled ECM-related protein complex are assayed. Data

WO 02/02772

PCT/US01/20729

5 obtained using different concentrations of ECM-related protein are used to calculate values for the number, affinity, and association of ECM-related protein with candidate ligand 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 invention. Although the invention has been described in connection with specific preferred 10 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. 15

WO 02/02772

PCT/US01/20729

TABLE 1

<u>Tissue</u>	<u>#cDNAs</u>	<u>Found</u>	<u>Abund</u>	<u>% Abund</u>
Cardiovascular System	270162	1/72	1	0.0004
Connective Tissue	147886	1/54	1	0.0007
Digestive System	514430	6/151	7	0.0014
Embryonic Structures	107325	2/23	2	0.0019
Endocrine System	233587	3/63	4	0.0017
Exocrine Glands	255105	4/64	5	0.0020
Genitalia, Female	445078	3/113	3	0.0007
Genitalia, Male	453150	5/118	7	0.0015
Germ Cells	46185	0/5	0	0.0000
Hemic and Immune System	701709	1/166	1	0.0001
Liver	110945	1/34	1	0.0009
Musculoskeletal System	162794	0/50	0	0.0000
Nervous System	973795	0/221	0	0.0000
Pancreas	111757	2/25	3	0.0027
Respiratory System	407942	7/95	9	0.0022
Sense Organs	25346	0/10	0	0.0000
Skin	72110	1/15	4	0.0055
Stomatognathic System	14025	1/11	1	0.0071
Unclassified/Mixed	150146	2/19	2	0.0013
Urinary Tract	287931	2/66	2	0.0007
Totals	5321883	33/1292	55	0.0000

WO 02/02772

PCT/US01/20729

TABLE 2

DE	P1Description	P2Description
1.7	Human, Colon, NrmL, mw/AdenoCA, Dn4614	Human, Colon Tumor, AdenoCA, Dn4614
2.3	Human, Colon Pool, NrmL, Dn3753	Human, Colon, Polyp, Dn3755
4.4	Human, Colon Pool, NrmL, Dn3753	Human, Colon Tumor, Cancer, Dn3311
1.7	Human, Colon Pool, NrmL, Dn3753	Human, Colon, Polyp, Dn3754
2.5	Human, Colon Pool, NrmL, mw/AdenoCA, Dn3583	Human, Colon Tumor, Adenoma, Dn3583
3.3	Human, Colon, NrmL, mw/AdenoCA, Dn3839	Human, Colon Tumor, AdenoCA, Dn3839
3.8	Human, Colon, Rectum, NrmL, mw/Cancer, Dn3581	Human, Colon Tumor, Rectum, Cancer, Dn3581
1.5	Human, Lung, NrmL, mw/Squamous Cell CA, Dn5796	Human, Lung Tumor, Squamous Cell CA, Dn5796
3.4	Human, Lung, NrmL, mw/Squamous Cell CA, Dn5800	Human, Lung Tumor, Squamous Cell CA, Dn5800

WO 02/02772

PCT/US01/20729

5 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.
2. An isolated cDNA comprising a nucleic acid sequence selected from:
 - 10 a) SEQ ID NO:2 or the complement thereof;
 - b) a fragment of SEQ ID NO:2 selected from SEQ ID NOs:3-13 or the complement thereof; and
 - c) a variant of SEQ ID NO:2 selected from SEQ ID NOs:14-15 or the complement thereof.
3. A composition comprising the cDNA of claim 1 and a labeling moiety.
- 15 4. A vector comprising the cDNA of claim 1.
5. A host cell comprising the vector of claim 4.
- 20 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.
7. A method for using a cDNA to detect expression of a nucleic acid in a sample comprising:
 - 25 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 of the cDNA in the sample.
- 30 8. The method of claim 7 further comprising amplifying the nucleic acids of the sample prior to hybridization.
9. The method of claim 7 wherein the composition is attached to a substrate.
- 35 10. The method of claim 7 wherein complex formation is compared with at least one standard to determine differential expression.
11. A method of using a cDNA to screen a plurality of molecules or compounds, the method

WO 02/02772

PCT/US01/20729

- 5 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.
- 10
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.
- 15
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;
 - b) an antigenic epitope of SEQ ID NO:1; and
 - c) a biologically active portion of SEQ ID NO:1.
- 20
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:
- 25
- 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
- 30
- 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:
- 35
- 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

WO 02/02772

PCT/US01/20729

- 5 the protein;
e) dissociating the antibodies from the protein, thereby obtaining purified antibodies.
18. An antibody produced by the method of claim 17.
- 10 19. A method for using an antibody to diagnose conditions or diseases associated with
expression of a protein, the method comprising:
a) combining the antibody of claim 18 with a sample, thereby forming antibody:protein
complexes; and
b) comparing complex formation with a standard, wherein the comparison indicates expression
15 of the protein in the sample.
20. The method of claim 19 wherein expression is diagnostic of a colon or lung cancer.
21. A pharmaceutical composition comprising the antibody of claim 18 and a pharmaceutical
20 carrier.
22. A method for treating a cancer comprising administering to a person in need of such
treatment an effective amount of the composition of claim 21.
- 25 23. The method of claim 22 wherein the cancer is a colon cancer.
24. The method of claim 22 wherein the cancer is a lung cancer.

WO 02/02772

PCT/US01/20729

1/13

```

5'  10  ATG  CTA  TTG  AGA  GTC  TTC  CTC  CAA  ACA  TCT  TCC  GAT  TTG  TTC  CTT  TAA  CCC  ATC  55
    19  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
    28  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
    37  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
    46  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
    64  TAG  ATC  TTC  GTC  GAA  ATC  AAT  TAC  AAA  CAT  TGC  CTT  ATG  TTG  GTT  TTC  TCG  AAC  109
    73  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
    82  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
    91  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   100  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   118  ACA  TTG  GCC  GAA  TAT  TGG  ATC  TTC  AGT  TGG  AGG  ACA  ACA  AAT  GGG  CCT  GCA  ATT  163
   127  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   136  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   145  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   154  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   172  GTG  ACT  TAT  TGC  AGT  TAA  AAA  CTT  GGT  TGG  AGA  ACA  TGC  CTC  CAC  AGT  CTA  TAA  217
   181  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   190  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   199  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   208  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   226  CTT  GGT  GAT  GTT  GTC  TGC  AAC  AGC  CCT  CCA  TTT  AAA  GGA  AGT  ATA  CTC  AGT  271
   235  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   244  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   253  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   262  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   280  AGA  CTA  AAG  AAG  GAA  TCT  AIT  TGC  CCT  ACT  CCA  CCA  GTG  TAT  GAA  GAA  CAT  GAG  325
   289  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   298  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   307  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   316  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   334  GAT  CCT  TCA  GGA  TCA  TTA  CAT  CTG  GCA  GCA  ACA  TCT  TCA  ATA  AAT  GAT  AGT  CGC  379
   343  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   352  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   361  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   370  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
   379  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---

```

FIGURE 1A

WO 02/02772

PCT/US01/20729

2/13

```

388  ATG TCA ACT AAG ACC ACG TCC ATT CTA AAA CTA CCC ACC AAA GCA CCA GGT TTG 433
-----
M  S  T  K  T  T  S  I  L  K  L  P  T  K  A  P  G  L
-----
442  ATA CCT TAT ATT ACA AAG CCA TCC ACT CAA CTT CCA GGA CCT TAC TGC CCT ATT 487
-----
I  P  Y  I  T  K  P  S  T  Q  L  P  G  P  Y  C  P  I
-----
496  CCT TGT AAC TGC AAA GTC CTA TCC CCA TCA GGA CTT CTA ATA CAT TGT CAG GAG 541
-----
P  C  N  C  K  V  L  S  P  S  G  L  L  I  H  C  Q  E
-----
550  CGC AAC ATT GAA AGC TTA TCA GAT CTG AGA CCT CCT CCG CAA AAT CCT AGA AAG 595
-----
R  N  I  E  S  L  S  D  L  R  P  P  Q  N  P  R  K
-----
604  CTC ATT CTA GCG GGA AAT ATT ATT CAC AGT TTA ATG AAG TCT GAT CTA GTG GAA 649
-----
L  I  L  A  G  N  I  I  H  S  L  M  K  S  D  L  V  E

```

FIGURE 1B

```

658   667   676   685   694   703
TAT TTC ACT TTG GAA ATG CTT CAC TTG GGA AAC AAT CGT ATT GAA GTT CTT GAA
Y F T L E M L H L G N N R I E V L E
-----
712   721   730   739   748   757
GAA GGA TCG TTT ATG AAC CTA ACG AGA TTA CAA AAA CTC TAT CTA AAT GGT AAC
E G S F M N L T R L Q K L Y L N G N
-----
766   775   784   793   802   811
CAC CTG ACC AAA TTA AGT AAA GGC ATG TTC CTT GGT CTC CAT AAT CTT GAA TAC
H L T K L S K G M F L G L H N L E Y
-----
820   829   838   847   856   865
TTA TAT CTT GAA TAC AAT GCC APT AAG GAA ATA CTG CCA GGA ACC TTT AAT CCA
L Y L E Y N A I K E I L P G T F N P
-----
874   883   892   901   910   919
ATG CCT AAA CTT AAA GTC CTG TAT TTA AAT AAC AAC CTC CTC CAA GTT TTA CCA
M P K L L K V L Y L N N N L L Q V L P
-----

```

FIGURE 1C

WO 02/02772

PCT/US01/20729

4/13

```

928  CCA CAT ATT TTT TCA GGG GTT CCT CTA ACT AAG GTA AAT CTT AAA ACA AAC CAG 973
-----
937  P H I F S G V P L T K V N L K T N Q 964
-----
982  TTT ACC CAT CTA CCT GTA AGT AAT ATT TTG GAT GAT CTT GAT TTA CTA ACC CAG 1027
-----
991  F T H L P V S N I L D D L D L L T Q 1018
-----
1036  ATT GAC CTT GAG GAT AAC CCC TGG GAC TGC TCC TGT GAC CTG GGT GGA CTG CAG 1081
-----
1045  I D L E D N P W D C S C D L V G L Q 1072
-----
1090  CAA TGG ATA CAA AAG TTA AGC AAG AAC ACA GTG ACA GAT GAC ATC CTC TGC ACT 1135
-----
1099  Q W I Q K L S K N T V T D D I L C T 1117
-----
1144  TCC CCC GGG CAT CTC GAC AAA AAG GAA TTG AAA GCC CTA AAT AGT GAA ATT CTC 1189
-----
1153  S P G H L D K K E L K A L N S E I L 1171
-----

```

FIGURE 1D

WO 02/02772

PCT/US01/20729

5/13

```

1198      1207      1216      1225      1234      1243
TGT CCA GGT TTA GTA AAT AAC CCA TCC ATG CCA ACA CAG ACT AGT TAC CTT ATG
-----
C P G L V N N P S M P T Q T S Y L M
-----
1252      1261      1270      1279      1288      1297
GTC ACC ACT CCT GCA ACA ACA AAT ACG GCT GAT ACT ATT TTA CGA TCT CTT
-----
V T T P A T T T N T A D T I L R S L
-----
1306      1315      1324      1333      1342      1351
ACG GAC GCT GTG CCA CTG TCT GTT CTA ATA TTG GGA CTT CTG ATT ATG TTC ATC
-----
T D A V P L S V L I L G L L I M F I
-----
1360      1369      1378      1387      1396      1405
ACT ATT GTT TTC TGT GCT GCA GGG ATA GTG GTT CTT GTT CTT CAC CGC AGG AGA
-----
T I V F C A A G I V V L V L H R R R
-----
1414      1423      1432      1441      1450      1459
AGA TAC AAA AAG AAA CAA GTA GAT GAG CAA ATG AGA GAC AAC AGT CCT GTG CAT
-----
R Y K K K Q V D E Q M R R D N S P V H
-----

```

FIGURE 1E

WO 02/02772

PCT/US01/20729

6/13

```

1468      1477      1486      1495      1504      1513
CTT CAG TTC ACC ATG TAT GGC CAT AAA ACC ACT CAT CAC ACT ACT GAA AGA CCC
-----
L  Q  Y  S  M  Y  G  H  K  T  T  H  H  T  T  E  R  P
-----
1522      1531      1540      1549      1558      1567
TCT GCC TCA CTC TAT GAA CAG CAC ATG GTG AGC CCC ATG GTT CAT CTC TAT AGA
-----
S  A  S  L  Y  E  Q  H  M  V  S  P  M  V  H  V  Y  R
-----
1576      1585      1594      1603      1612      1621
AGT CCA TCC TTT GGT CCA AAG CAT CTG GAA GAG GAA GAA GAG AGG AAT GAG AAA
-----
S  P  S  F  G  P  K  H  L  E  E  E  E  E  R  N  E  K
-----
1630      1639      1648      1657      1666      1675
GAA GGA AGT GAT CCA AAA CAT CTC CAA AGA AGT CTT TTG GAA CAG GAA AAT CAT
-----
E  G  S  D  A  K  H  L  Q  R  S  L  L  E  Q  E  N
-----
1684      1693      1702      1711      1720      1729
TCA CCA CTC ACA GGG TCA AAT ATG AAA TAC AAA ACC ACG AAC CAA TCA ACA GAA
-----
S  P  L  T  G  S  N  M  K  Y  K  T  T  N  Q  S  T  E

```

FIGURE 1F

WO 02/02772

PCT/US01/20729

7/13

```

1738      1747      1756      1765      1774      1783
TTT TTA TTC CAA GAT GCC AGC TCA TTG TAC AGA AAC ATT TTA GAA AAA GAA
-----
F L S F Q D A S S L Y R N I L E K E
1792      1801      1810      1819      1828      1837
AGG GAA CTT CAG CAA CTG GGA ATC ACA GAA TAC CTA AGG AAA AAC ATT GCT CAG
-----
R E L Q Q L G I T E Y L R K N I A Q
1846      1855      1864      1873      1882      1891
CTC CAG CCT GAT ATG GAG GCA CAT TAT CCT GGA GCC CAC GAA GAG CTG AAG TTA
-----
L Q P D M E A H Y P G A H E E L K L
1900      1909      1918      1927      1936      1945
ATG GAA ACA TTA ATG TAC TCA CGT CCA AGG AAG GTA TTA GTG GAA CAG ACA AAA
-----
M E T L M Y S R P R K V L V E Q T K
1954      1963      1972      1981      1990      1999
AAT GAG TAT TTT GAA CTT AAA GCT AAT TTA CAT GCT GAA CCT GAC TAT TTA GAA
-----
N E Y F E L K A N L H A E P D Y L E

```

FIGURE 1G

WO 02/02772

PCT/US01/20729

8/13

```

2008      2017      2026      2035      2044      2053
GTC CTG GAG CAG CAA ACA TAG ATG GAG AGT TTG AGG GCT TTC GCA GAA ATG CTG
-----
V L E Q Q T
2062      2071      2080      2089      2098      2107
TGA TTC TGT TTT AAG TCC ATA CCT TGT AAA TAA GTG CCT TAC GTG AGT GTG TCA
-----
2116      2125      2134      2143      2152      2161
TCA ATC AGA ACC TAA GCA CAG CAG TAA ACT ATG GGG AAA AAA AAA GAA GAA GAA
-----
2170      2179      2188      2197      2206      2215
AAA GAA ACT CAG GGA TCA CTG GGA GAA GCC ATG GCA TTA TCT TCA GGC AAT TTA
-----
2224      2233      2242      2251      2260      2269
GTC TGT CCC AAA TAA AAT AAA TCC TTG CAT GTA AAT CAT TCA AGG GTT ATA GTA
-----
2278      2287      2296      2305      2314      2323
ATA TTT CAT ATA CTG AAA AGT GTC TCA TAG GAG TCC TCT TGC ACA TCT AAA AAG
-----

```

FIGURE 1H

WO 02/02772

PCT/US01/20729

9/13

```

2332 2341 2350 2359 2368 2377
GCT GAA CAT TTA AGT ATC CCG AAT TTT CTT GAA TTG CTT TCC CTA TAG ATT AAT
-----
2386 2395 2404 2413 2422 2431
TAC AAT TGG AIT TCA TCA TTT AAA AAC CAT ACT TGT AFA TGT AGT TAT AAT ATG
-----
2440 2449 2458 2467 2476 2485
TAA GGA ATA CAT TGT TTA TAA CCA GTA TGT ACT TCA AAA ATG TGT ATT GTC AAA
-----
2494 2503 2512 2521 2530 2539
CAT ACC TAA CTT TCT TGC AAT AAA TGC AAA AGA AAC TGG AAC TTG ACA ATT ATA
-----
2548 2557 2566 2575 2584 2593
AAT AGT AAT AGT GAA GAA AAA ATA GAA AGG TTG CAA TTA TAT AGG CCA TGG GTG
-----
2602 2611 2620 2629 2638 2647
GCT CAA AAC TTT GAA CAT TTG AGC TTA AAC AAA TGC CAC TCT CAT GCA TTC TAA
-----

```

FIGURE 1I

WO 02/02772

PCT/US01/20729

10/13

```

2656 2665 2674 2683 2692 2701
ATT AAA TTA AAA TGA TTA ATA GTT CAG GTG GAA GAA ATA AGC ATA CTT TTT
-----
2710 2719 2728 2737 2746 2755
GGG TTT TCT ACA CAT TTT GTG TAG ACA ATT TTA ATG TCA GTG CTG CTG TGA ACT
-----
2764 2773 2782 2791 2800 2809
AAA GTA TGT CAT TTA TGC TCA AAG TTT AAT TCT TCT TGG GAT ATT TTA AAA
-----
2818 2827 2836 2845 2854 2863
ATG CTA CTG AGA TTC TGC TGT AAA TAT GAC TAG AGA ATA TAT TGG GTT TGC TTT
-----
2872 2881 2890 2899 2908 2917
ATT TCA TAG GCT TAA TTC TTT GTA AAT CTG AAT GAC CAT AAT AGA AAT ACA TTT
-----
2926 2935 2944 2953 2962 2971
CTT GTG GCA AGT AAT TCA CAG TTG TAA AGT AAA TAG GAA AAA TTA TTT TAT TTT
-----
2980 2989 2998 3007 3016 3025
TAT TGA TGT ACA TTG ATA GAT GCC ATA AAT CAG TAG CAA AAG GCA CTT CTA AAG
-----

```

FIGURE 1J

WO 02/02772

PCT/US01/20729

11/13

```

3034 3043 3052 3061 3070 3079
GTA AGT GGT TTTA AGT TGC CTC AAG AGA GGG ACA ATG TAG CTT TAT TTT ACA AGA
-----
3088 3097 3106 3115 3124 3133
AGG CAT AGT TAG ATT TCT ATG AAA TAT TTA TTC TGT ACA GTT TTA TAT ATT TTT
-----
3142 3151 3160 3169 3178 3187
GGT TCA CAA AAG TAA TTA TTC TTG GGT GCC TTT CAA GAA AAT TAA AAA TAC TAC
-----
3196 3205 3214 3223
CCA CTA CAA TAA AAC TAA AAT GAA AAC TCA AAA AAA A 3'
-----

```

FIGURE IK

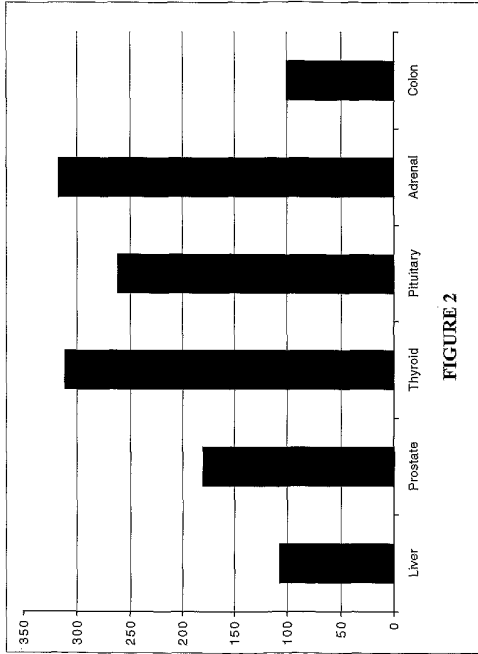


FIGURE 2

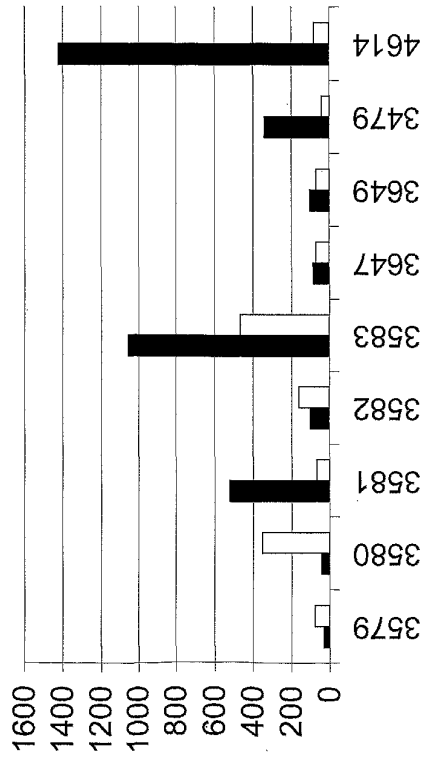


FIGURE 3

WO 02/02772

PCT/US01/20729

<110> INCYTE GENOMICS, INC.
LASEK, Amy K. W.
YUE, Henry
PATTERSON, Chandra

<120> BCM-RELATED TUMOR MARKER

<130> PC-0045 US

<140> To Be Assigned

<141> Herewith

<150> 60/215,454

<151> 2000-06-30

<160> 15

<170> PERL Program

<210> 1

<211> 546

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2743093CD1

<400> 1

```

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

```

WO 02/02772

PCT/US01/20729

His Leu Asp Lys Lys Glu Leu Lys Ala Leu Asn Ser Glu Ile Leu
 260 265 270
 Cys Pro Gly Leu Val Asn Asn Pro Ser Met Pro Thr Gln Thr Ser
 275 280 285
 Tyr Leu Met Val Thr Pro Ala Thr Thr Thr Asn Thr Ala Asp
 290 295 300
 Thr Ile Leu Arg Ser Leu Thr Asp Ala Val Pro Leu Ser Val Leu
 305 310 315
 Ile Leu Gly Leu Leu Ile Met Phe Ile Thr Ile Val Phe Cys Ala
 320 325 330
 Ala Gly Ile Val Val Leu Val Leu His Arg Arg Arg Arg Tyr Lys
 335 340 345
 Lys Lys Gln Val Asp Glu Gln Met Arg Asp Asn Ser Pro Val His
 350 355 360
 Leu Gln Tyr Ser Met Tyr Gly His Lys Thr Thr His His Thr Thr
 365 370 375
 Glu Arg Pro Ser Ala Ser Leu Tyr Glu Gln His Met Val Ser Pro
 380 385 390
 Met Val His Val Tyr Arg Ser Pro Ser Phe Gly Pro Lys His Leu
 395 400 405
 Glu Glu Glu Glu Glu Arg Asn Glu Lys Glu Gly Ser Asp Ala Lys
 410 415 420
 His Leu Gln Arg Ser Leu Leu Glu Gln Glu Asn His Ser Pro Leu
 425 430 435
 Thr Gly Ser Asn Met Lys Tyr Lys Thr Thr Asn Gln Ser Thr Glu
 440 445 450
 Phe Leu Ser Phe Gln Asp Ala Ser Ser Leu Tyr Arg Asn Ile Leu
 455 460 465
 Glu Lys Glu Arg Glu Leu Gln Gln Leu Gly Ile Thr Glu Tyr Leu
 470 475 480
 Arg Lys Asn Ile Ala Gln Leu Gln Pro Asp Met Glu Ala His Tyr
 485 490 495
 Pro Gly Ala His Glu Glu Leu Lys Leu Met Glu Thr Leu Met Tyr
 500 505 510
 Ser Arg Pro Arg Lys Val Leu Val Glu Gln Thr Lys Asn Glu Tyr
 515 520 525
 Phe Glu Leu Lys Ala Asn Leu His Ala Glu Pro Asp Tyr Leu Glu
 530 535 540
 Val Leu Glu Gln Gln Thr
 545

<210> 2

<211> 3224

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2743093CB1

<400> 2

gatgctattg agagtcttcc tccaaacatc ttccgatttg ttcccttaac ceatctagat 60
 cttcgtggaa atcaatbaca aacattgcct tatgttgggt ttctcgaaca cattggccga 120
 atattggatc ttcagttgga ggacaacaaa tggcoctgca attgtgactt attgcagtta 180
 aaaacttggc tgggagaacat gcctccacag tctataaactt ggtgatgttg tctgcaacag 240
 ccctccattt tttaaaggaa gtataactcag tagactaaag aaggaatcta tttyccctac 300
 tccaccagtg tatgaagaac atgaggatcc ttcagatca ttacacttg cagcaacac 360
 ttcaataaat gatagtcgca tgtcnaactaa gaccacgtcc attctaaaac taccaccaa 420
 agcaccaggt ttgatacctt atattacaaa gccatccact caacttccag gaccttactg 480
 ccctattcct tgbactgca aagctctatc cccatccagg cttctaatac attgtcagga 540
 gcgcacattt gaagcttat cagatctgag acctctctcg caaaatccta gaaagctcat 600
 tctagcggga aatattatc acagtttaac gaagtctgat ctagtggat atttcacttt 660
 ggaatgctt cacttgggaa acaatcgtat tgaagttcct gaagaaggat cgtttatgaa 720

WO 02/02772

PCT/US01/20729

```

cctaacgaga ttacaaaaac tctatctaaa tggtaaccac ctgaccaaat taagtaaaag 780
catgttccctt ggtctccata atcttgaata cttatatctt gaatacaatg ccattaagga 840
aatactgcca ggaaccttta atccaatgcc taaacttaaa gtccctgtatt taaataaaca 900
octctccaaa gttttaccac cacatatctt ttcaggggtt cctctaaact aggtaaactt 960
taaaaacaaac cagtttaccct acctaccctgt aagtaataat ttggatgctc ttgattactt 1020
aaoccaatgt gaccttgagg ataacccttg ggaactgtcc ttggaacttg ttggactgca 1080
gcaatggata caaaaagttaa gcaagaacac agtgacagat gaactcctct gcaactcccc 1140
cgggcatctc gacaaaaggg aattgaaagc cctaataagt gaaattctct gtccaggttt 1200
agtaataaac ccatccatgc caacacagac tagttacctt atggtcacca ctccctgcaac 1260
aacaacaat agcgctgata ctatcttacg atctcttacg gacgctgtgc caactgtctgt 1320
tctaataatg ggaactctga ttatgttcat cactattgtt tctgtgctg cagggatagt 1380
ggttcttgtt ctaccocgca ggagaagata caaaaagaaa caagttagat agcaaatgag 1440
agacaaagt cctgtgctc ttcaatagc catgtatggc cataaaacca ctcaatcacac 1500
tactgaaaga ccctctgctt cactctatga acagcacatg gtgagcccca tggttcatgt 1560
ctatagaagt ccactccttg gtccaaagca tctggagag gaagaagaga ggaatgaga 1620
agaaggaaat gatgcaaaac atctccaaag aagtcttttg gaacaggaaa atccattacc 1680
actccacagg tcaaaatga aatacaaaac cacgaaccaa tcaacagaat ttttatccct 1740
ccagatgccc agctcattgt acagaaacat tttgaaaaa gaaggggaac ttccagcaact 1800
gggaatcaca gaatacctaa ggaacaaat tgcctagctc cagcctgata tggaggcaca 1860
ttatcctgga gccacggaag agctgaagtt aatggaaca ttaatgtact caactccag 1920
gsaggatata gtggaacaga caaaaatga gtatcttgaa cttaagacta atttaccatg 1980
tgaacctgac tattagaag tctctggagc gaaacatag atggaggtt tggaggcttt 2040
cgagaaatg ctgtgatctt gttttaaact catacctgt aataaagtc cttactgtgag 2100
tgtgtcatca atcagaacct aagcacagca gtaactatg gggaaaaaa aagaagaaga 2160
aaaaaact cagggatcac tgggagaagc catggcatta tcttcagcca atttagctg 2220
tcccaataaa aataaatcct tgcagttaaa tcatccaagg gtatatgtaa tattcatat 2280
actgaaagt gctcataggt agtccctctg cacatctaaa aaggctgaac atttaatgat 2340
ccgaaatctt ctgaaatgc tttccctata gattaattac aattggattt catcatttaa 2400
aaacataact tgtatagtga gttataatat gtaaggaata cattgttat aaccagtatg 2460
taattcaaaa atgtgtattg tcaaacatgc ctaactttct tgcataaat gcaaaagaaa 2520
ctggaactgt acaatctaaa atagtaatg tggagaaaa atagaaagg tgcattata 2580
tagccatgg gttgctcaaa actttgaaac tttgactta acaaatgctt actctcatgc 2640
atctcaaat aaaaagttaa atgattaat agttcaggtg gaagaaataa gcatactttt 2700
tgggtttctt acacattttg tbtagcaact ttaatgtca gtgctgtgt gaaactaaat 2760
atgtcattta tgcctaaagt ttaattctt tcttgggat attttaaaa tgcactagac 2820
attctgctgt aatatgact agagaatata tgggtttgc tttattcat aggcctaat 2880
ctttgtaaat ctgaatgacc ataatagaaa tacatttctt tgggcaagta attcacagtt 2940
gtaagtaaaa taggaaaaat tattttattt ttattgatgt acattgatag atgccataaa 3000
tcagtagcaa aaggcacttc taaaggtaag tggtttaagt tgcctcaaga gaggacaaat 3060
gtagcttat ttacaagaa ggcatagtta gatttctatg aatatittat tctgtacagt 3120
ttatatatt tttggtcac aaaaataatt attcttgggt gcctttcaag aaaattaaaa 3180
atactacca ctacaataaa actaaaatga aaactcaaaa aaaa 3224

```

```

<210> 3
<211> 246
<212> DNA
<213> Homo sapiens
<220>
<221> misc_feature
<223> Incyte ID No: 2743093H1

```

```

<400> 3
tgcttcactt gggaaacaat cgtattgaag ttcttgaaga aggatcgttt atgaacctaa 60
cgagattaca aaaactctat ctaaatggta accacctgac caaattaagt aaaggcatgt 120
tccttggctt ccataatctt gaatacttat atcttgaata caatgccatt aaggaaatcc 180
tgcaggaaac ctttaatcca atgocataaac ttaaagtctt gtatttaaat acaaacctcc 240
tccaag 246
<210> 4
<211> 597
<212> DNA
<213> Homo sapiens

```

WO 02/02772

PCT/US01/20729

<220>
<221> misc_feature
<223> Incyte ID No: 4876623F9

<400> 4
cctattcctt gtaactgcaa agtcctatcc ccatcaggac ttctaataca ttgtcaggag 60
cgcacacatg aaagcctatc agatctgaga cctctccgc aaatctctg aaagctcatt 120
ctagcgggaa atattattca cagtttaag aagctcgatc tagtggaaata tttcactttg 180
gaatgccttc acttgggaaa caatcgtatt gaagtctctg aagaaggatc gtttatgaac 240
ctaacgagat tacaanaact ctatctaact ggtanccacc tgaccaaat aagtaaggc 300
atgttctctg gctccataa tcttgaatac ttatatcttg aatacaatgc cattaaggaa 360
atactgccag gaacctttaa tcaatgcct aaacttaaag tctgtatctt aaataacaac 420
ctctccaag ttttaaccacc acatattttt tcaggggttc ctcataacta aggtaaatct 480
taaaccaaac cagtttacc atctacctgt aagtaaatct tctggatgat cttgatttac 540
taaccagat tgacctgag gataaccctt gggactgctc ctgtgacctg gttggac 597

<210> 510
<211> 510
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2316239T6

<220>
<221> unsure
<222> 484
<223> a, t, c, g, or other

<400> 5
tccccatagt ttactgctgt gcttaggttc tgattgatga cacactcacg taaggcaact 60
atttacaag tatggactta asacagaatc acagcatttc tgcgaagcc ctcaactct 120
ccatctatgt ttgctgctcc aggacttcta aatagtcagg ttcagcatgt aaattagctt 180
taagtccaaa ataactcattt ttgtctgttt ccaactaatc ctctcttggc cgtgagtaca 240
ttaagtcttc cattaacttc agctctctgt gggctccagc ataagtgcc tccatctcag 300
gctggagctg agcaatgttt ttccttaggt attctgtgat tcccagttgc tgaagtccc 360
ttctcttttc taaaatgttt ctgtacaatg agctggcacc ttggaggatc aaaaattctg 420
ttgattgggt cgtgggtttg tattcatat ttgacctgt gagtggtgaa tgattttctc 480
gttcaaaaag acttcttttg agatgtttg 510

<210> 6
<211> 549
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6258015F8

<400> 6
aaaaacttga ggaggttgtt atttaaatac aggactttaa gtttaggcat tggattaaag 60
gttcttggca gtatttcttt aatggcattg tattcaagat ataagtattc aagattatgg 120
agacctaggc acatgccttt acttaatttg gtcaggtggt taccatttag atagagtttt 180
tghtaatctg alttaggttca taaacgatcc ttcttcaaga acttcaatc gattgtttcc 240
caagtgaagc atttccaaag tgaatattc cactagatca gactcaata aactgtgaat 300
aaatctccc gctagaatga gctttctagg attttgcgga ggaagctcct gatctgataa 360
gctttcaatg ttgcgtctct gacaatgat tsgaagtctt gatgggata ggaactttga 420
gttacaagga atagggcagt aaggtctctg aagttgagtg gatgctttg taatataagg 480
tatcaaacct ggtgcttttg tgggtagttt tagaatggac gtggtcttag ttgacatgca 540
actatcatt 549

WO 02/02772

PCT/US01/20729

```

<210> 7
<211> 466
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7677606J1

<400> 7
ggaagttgag tggatggctt agtaataata ggtatcaaac ctggtgcttt ggtgggtagt 60
tttagaatgg acgtggtctt agttgacatg cgactatcat ttattgaaga tgttgctgoc 120
agalgtaatg atcctgaagg atcctcatgt tcttcataca ctggtggagt agggcaataa 180
gattcctctt ttagtctact gagtatactt ccttataaaa atggagggtt gttgcagaca 240
acatcaccaa gttatagact ggggagcatg gttctccaac caagttttta actgcaataa 300
gtcaccaatt caggcccatt gttgtctctc caactgaaga tccaatattc ggccaatglg 360
ttcagaaaaa ccaacataag gcaatgtttg taatgtactt ccacgaagt ctgatgggt 420
taaaggaaaa aatcgaaga tgtttggagg aagactctca atagca 466

<210> 8
<211> 640
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71111915V1

<400> 8
aatcttgaat acttataatc tgaatacaat gccattaagg aaatactgcc aggaaccttt 60
aatccaatgc ctaaaactaa agtctctgat ttaataaaca acctcctcca agttttacca 120
ccacatattt ttccaggcgt tctctaaact aaggtaaatc ttaaaaacaa ccagtttacc 180
catctacctg taagtaatat ttggatgat cttgatttac taaccagatg tgaccttgag 240
gataaccctc gggactgctc ctgtgacctg gttggactgc agcaatggat acaaaagtta 300
agcaagaaca cagtgcacaga tgacatcttc tgcacttccc cgggcatct cgacaaaag 360
gaattgaaag ccctaaaatg tgaatctctc tgtccaggtt tagtaataaa cccatccatg 420
ccaacacaga ctagttaact tatggtcacc actcctgcaa caacaacaaa tacggctgat 480
actattttac gatctcttac ggaagctgtg ccactgtctg ttctaataat gggacttctg 540
attatgttca tcaactattg tttctgtgct gcagggatag tggttcttgt tcttcaccgc 600
aggagaagat acaaaaagaaa caagtagatg agcaatgaga 640

<210> 9
<211> 638
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71112850V1

<400> 9
caabatattc tctagtcata tttacagcag aatctcagta gcaattttaa aatatcccaa 60
gaagaagaat taactttga gcataaatga catactttag ttccacagca gcaactgacat 120
taaaatgtc tacacaaaat gtgtagaaaa cccaaaaagt atgcttattt ctccaccctg 180
aaactataat catttlaact ttttaaltta gaatgcataa gagtggcatt tgttlaagct 240
caaatgttca agttttgag ccacctagg cctatataat tgcacacttt ctatttttct 300
ttcaactata ctattataa ttgtcaagtt ccagtttctt ttgcatttat tgcagaaga 360
ttaggtatgt ttgacaatac acatltttga agtacatact ggtataaac aatgtatctc 420
tcaatatta taactacata tacaagtatg gtttttaaat gatgaaatcc aatgttaatt 480
aatctatagg gaagcaatt caagaaatlt cgggatactt aatgttccg cctltttaga 540
tgtgcaagag gactcctatg agacactttt cagtatatga aatattacta taaccttgaa 600
tgatttacat gcgagatcta ttttattggg acagacta 638

```

WO 02/02772

PCT/US01/20729

<210> 10
<211> 582
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71262960V1

<400> 10
gcagcaatgg tatacaaanag ttaagcaaga acacagtgac agatgacatc ctctgcactt 60
ccccgggca tctcgacaaa aaggaattga aagccctaaa tagtgaatt ctctgtccag 120
gttttagtaaa taacctatcc atgccaacac agactagtta ccttatggtc accactcctg 180
caacaacaac aaatcaggct gatactatct tacgatctct tacggagcgt gtgcactctg 240
ctgttctaata altgggactt ctgattatgt tcatcactat tgttttctgt gctgcaggga 300
tagtgggtctc tgttcttccac cgcaggagaa gatacaaaaa gaaacaagta gatgagcaaa 360
tgagagcaaa cagtcctctg catcttcagt acagcatgta tggccataaa accactcact 420
acactactga agacccctct gctcactct atgaaacaga catgggtagc cccatgggttc 480
atgtctatag aagtcacatc ttgggtccaa agcatctgga agaggaagaa gagaggaatg 540
agaagaagg aagtgtgca aaacatctcc aaagaagtct tt 582

<210> 11
<211> 494
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71264035V1

<400> 11
ttttgtgtag acaattttaa tgtcagtgct gctgtgeact aagatgtgc atttatgctc 60
aaagttaaat tcttcttctt gggatatttt aaaaatgcta ctgagattct gctgtaata 120
tgactagaga atatatggg ttgcttttat tctcaggct taattctttg taaatctgaa 180
tgaccataat agaatacat tcttctgggc aagtaattca cagttgtaa gtaaatagga 240
aaaattattt tatttttatt gatgtacatt gatagatgcc ataatcagt agcaaaagcc 300
acttcaaaag gtaagtgttt taagttgcct caagagaggg acaatgtagc tttattttac 360
aagaaggcat agttagattt ctatgaata tttattctgt acagttttat atatttttgg 420
ttcaaaaag taattattct tgggtgcctt tcaagaaaat taaaaatact accactaca 480
ataaaactaa aatg 494

<210> 12
<211> 444
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71113484V1

<400> 12
taagcctatg aaataaagca aacccaatat attctctagt catatttaca gcagaatctc 60
agtagcattt taaaaatctc ccaagaagaa gaattaaact ttgagcataa atgacatact 120
ttagtctcac agcagcactg acattaaaat gctacacaaa aatgtgtaga aaacccaaaa 180
agtatgttta ttcttccac ctgaactact aatcatttta actttttaa ttagaatgca 240
tgagagggc atttgtttaa gctcaatgt tcaagtttt gggccacca tggcctatat 300
aattgcaccc ttctcattt ttcttccacta ttactattha taattgtcaa gttccagttt 360
cttttgcatc tattgcaaga aagtbaggta tgtttgacaa tacacatttt tgaagtacat 420
actggttata aacaatgtat tctt 444

<210> 13
<211> 657
<212> DNA

WO 02/02772

PCT/US01/20729

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 71114738V1

<400> 13

```

caggctcaat atgaaataca aaaccacgaa ccaatcaaca gaatttttat ccttccaaga 60
tgccagctca ttgtacagaa acattttaga aaaagaagg gaacttcagc aactgggaat 120
cacagaatac ctaaggaaaa acattgctca gctccagcct gatatggagg cacattatcc 180
tggagcccaac gaagagctga agttaatgga aacattaatg tactcacgtc caaggaaggt 240
attagtgaaa cagacaaaaa atgagtattt tgaacttaaa gctaatttac atgctgaacc 300
tgactattta gaagtcctgg agcagcaaac atagatggag agtttgaggg ctttcgcaga 360
aatgtctgta ttctgtttta agtccatacc ttgtaataaa gtgccttacg tgagtgtgtc 420
atcaatcaga acctaaagcac agcagtaaac tatggggaac aaaaaaagaa gaagaaaaga 480
aactcagga tcactgggag aagccatggc attatcttca gcaaatatag tctgtcccaa 540
ataaaataaa tccttgcatg taatcattc aggttatag taatatttca talactgaaa 600
agtgctcat agggagtctt cttgcacatc taaaaggct gacatttaag tatccga 657

```

<210> 14

<211> 610

<212> DNA

<213> Canis familiaris

<220>

<221> misc_feature

<223> Incyte ID No: 703545995J1

<400> 14

```

caaccaaatt tttagctcta ataagtcaca attacaggtc cacttatgtc ctccaactgg 60
agatccaata tccggccaat gtgttctaag aaaccaactg aaggcaatgt tgcaactgat 120
ttcccagaag atctagatgg gttaaaggaa caatcgaaa tatgtttgga ggaagactct 180
caatagcatt gtcatttaaa ataacactt taagtctgtt gagcttgcta aagcacttg 240
gttcaatcac tgtaatgaag ttgttgtctg cttgtaggaa ttctaggttt tccagctcat 300
ggaaggtatc ctcttlaaga atttctaag aatttgatt gatgtgaagt tgcttaagaa 360
ggccaagggc attaatgca ccagctcaca tatctgcaat gttgtlaaat ccaaggggt 420
aatgagatgg cattagtaag ccagaaaaag tcatttgggt gaagcattgt caaacatta 480
tttaataaac ttaggtgaaa aggtcgtgat ggcggcacac ttatttggga taactcttg 540
ataccttctt cttcacagtt tattagcatt gtgccatctt ttttcctcac aattgcaaa 600
agaatcacag

```

<210> 15

<211> 273

<212> DNA

<213> Mus musculus

<220>

<221> misc_feature

<223> Incyte ID No: 112357_Mm.1

<400> 15

```

tgtgaagtc attgttcaac thattttttt ttcattgaca taccctgogt aggtcgteta 60
gcttagtagc aagaggogtt tctacaggtt agtgttttga gcttcctcag gagagggaca 120
gtgtagcttt atttcaag aaggtatcac tagatttita tgaattattt attttgtaca 180
gtttgtata ttttggctc acaagataat tatacttggg tgccttttaa gaaagttta 240
gattattgct cacttcaata aaagtaaaat gaa

```

【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
10 January 2002 (10.01.2002)

PCT

(10) International Publication Number
WO 02/002772 A3

(51) International Patent Classification: C12N 15/12, 94087 (US), PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US).
C07K 14/78, 16/18, C12Q 1/68, G01N 33/53, A61K 38/17, 39/395, A61P 35/00

(21) International Application Number: PCT/US01/20729

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

(22) International Filing Date: 29 June 2001 (29.06.2001)

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, IE, 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.

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data: 60/215,454 30 June 2000 (30.06.2000) US

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

(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, FI, FR, GB, GR, IE, IT, LI, MC, NL, PT, SE, TR), OAPI patent (BF, BI, CF, CG, CI, CM, GA, GN, GW, ML, MR, NI, SN, TD, TG).

(72) Inventors; and
(75) Inventors/Applicants (for US only): LASEK, Amy, K., W. [US/US]; 237 41st Street, #5, Oakland, CA 94611 (US);
YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA

Published:
with international search report

[Continued on next page]

(54) Title: HUMAN EXTRACELLULAR MATRIX (ECM)-RELATED TUMOR MARKER

5' ATG CTA TGG AGA GTC TTC CTC CAA ACA TCT TCC GAT TGG TTC CTT TAA CCG ATC

64 TAG ATC TTC GTG GAA ATC AAT TAC AAA CAT TGC CTT ACG TGG GTT TTC TCG AAC

118 ACA TTG GCC GAA TAT TGG ATC TTC AGT TGG AGG ACA ACA AAT GGG CCT GCA AAT

172 CTG ACT TAT TGC AGT TAA AAA CTT GGT TGG AGA ACA TGC CTC CAC AGT CTA TAA

226 CTT GST GAT GTT CTC TGC AAC AGC CCT CCA TTT TTT AAA GGA AGT ATA CTC AGT

280 AGA CTA AAG AAG GAA TCT ATT TGC CCT ACT CCA CCA CTG TTT GAA GAA CAT GAG

334 GAT CCT TCA GGA TCA TTA CAT CTG GCA GCA ACA TCT TCA ATA AAT GAT AGT CCG



WO 02/002772 A3

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

WO 02/002772 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		International Application No. PCT/US 01/20729
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/12 C07K14/78 C07K16/18 C12Q1/68 G01N33/53 A61K38/17 A61K39/395 A61P35/00		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 7 C07K C12N		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) SEQUENCE SEARCH, EPO-internal, WPI Data, PAJ, BIOSIS		
C. 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! 27 January 2000 (2000-01-27) OTTENWALDER ET AL.: "Homo sapiens mRNA; cDNA DKFZp564o1278" Database accession no. AL137517 XP002214969 Comprises at least SEQ ID NO:3 of the application and encodes part of SEQ ID NO:1 abstract	2, 13, 15-19
E	WO 01 81363 A (MURDOCK PAUL R ;SMITHKLINE BEECHAM PLC (GB); SMITHKLINE BEECHAM CO) 1 November 2001 (2001-11-01) the whole document See polypeptide #10 (SEQ ID NO:49)	1-24
<input type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents: *A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. *Z* document member of the same patent family		
Date of the actual completion of the international search 27 September 2002		Date of mailing of the international search report 11/10/2002
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3916		Authorized officer Sprinks, M

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 01/20729

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. <input checked="" type="checkbox"/>	<p>Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely: Although claims 22-24 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.</p>
2. <input type="checkbox"/>	<p>Claims Nos.: because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:</p>
3. <input type="checkbox"/>	<p>Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).</p>
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. <input type="checkbox"/>	As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. <input type="checkbox"/>	As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. <input type="checkbox"/>	As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. <input type="checkbox"/>	No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims, it is covered by claims Nos.:
Remark on Protest	<input type="checkbox"/> The additional search fees were accompanied by the applicant's protest. <input type="checkbox"/> No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT				International Application No PCT/US 01/20729	
Patent document cited in search report	Publication date	Patent family member(s)		Publication date	
WO 0181363	A	01-11-2001	AU WO	5726501 A 0181363 A1	07-11-2001 01-11-2001

フロントページの続き

(51) Int.Cl. ⁷	F I	テーマコード(参考)
A 6 1 K 48/00	A 6 1 K 48/00	4 C 0 8 4
A 6 1 P 35/00	A 6 1 P 35/00	4 C 0 8 5
C 0 7 K 1/22	C 0 7 K 1/22	4 C 0 8 6
C 0 7 K 14/47	C 0 7 K 14/47	4 H 0 4 5
C 0 7 K 16/18	C 0 7 K 16/18	
C 1 2 N 1/15	C 1 2 N 1/15	
C 1 2 N 1/19	C 1 2 N 1/19	
C 1 2 N 1/21	C 1 2 N 1/21	
C 1 2 N 5/10	C 1 2 P 21/08	
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	M
G 0 1 N 33/53	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, 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, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW

(72) 発明者 ユエ、ヘンリー

アメリカ合衆国カリフォルニア州 9 4 0 8 7 ・ サニーベイル・ルイスアベニュー 8 2 6

(72) 発明者 アービズ、チャンドラ・エス

アメリカ合衆国カリフォルニア州 9 5 1 2 5 ・ サンノゼ・モロッコドライブ 1 7 0 6

F ターム(参考) 2G045 AA35 BB10 BB20 BB29 BB46 BB50 CB01 CB02 DA13 FA12
 FB02 FB12 GC15
 4B024 AA01 AA11 BA80 CA04 DA02 DA05 DA06 DA11 DA12 EA04
 GA11 HA12
 4B063 QA18 QA19 QQ03 QQ08 QQ43 QR08 QR42 QR56 QS25 QS34
 QX02 QX07
 4B064 AG26 CA10 CA20 CC24 DA01 DA13
 4B065 AA01X AA57X AA72X AA90X AA93Y AB01 BA02 CA24 CA44 CA46
 4C084 AA02 AA06 AA07 AA13 AA17 BA01 BA08 BA22 BA23 CA17
 MA01 NA14 ZB212 ZB262
 4C085 AA13 AA14 CC32 EE01 GG01
 4C086 AA01 AA02 AA03 AA04 EA16 MA01 MA04 NA14 ZB21 ZB26
 4H045 AA10 AA11 AA20 AA30 BA10 CA40 DA75 EA28 EA51 FA72
 FA74

专利名称(译)	相关肿瘤标志物		
公开(公告)号	JP2004516813A	公开(公告)日	2004-06-10
申请号	JP2002508012	申请日	2001-06-29
[标]申请(专利权)人(译)	洞察Genomics公司		
申请(专利权)人(译)	洞察基因组公司		
[标]发明人	ラセクアミーケイダブリュ ユエヘンリー アービズチャンドラエス		
发明人	ラセク、アミーケイダブリュ ユエ、ヘンリー アービズ、チャンドラエス		
IPC分类号	G01N33/50 A61K31/7088 A61K38/00 A61K39/395 A61K45/00 A61K48/00 A61P35/00 C07K1/22 C07K14/47 C07K14/78 C07K16/18 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N15/09 C12N15/12 C12P21/08 C12Q1/68 C12Q1/6883 G01N33/15 G01N33/53 G01N33/566 G01N33/574		
CPC分类号	A61K38/00 A61P35/00 C07K14/78 C12Q1/6883 C12Q2600/158 G01N33/57419 G01N33/57423 G01N2500/00		
FI分类号	C12N15/00.ZNA.A A61K31/7088 A61K39/395.E A61K39/395.T A61K45/00 A61K48/00 A61P35/00 C07K1/22 C07K14/47 C07K16/18 C12N1/15 C12N1/19 C12N1/21 C12P21/08 C12Q1/68.A G01N33/15. Z G01N33/50.Z G01N33/53.M G01N33/566 C12N5/00.A A61K37/02		
F-TERM分类号	2G045/AA35 2G045/BB10 2G045/BB20 2G045/BB29 2G045/BB46 2G045/BB50 2G045/CB01 2G045 /CB02 2G045/DA13 2G045/FA12 2G045/FB02 2G045/FB12 2G045/GC15 4B024/AA01 4B024/AA11 4B024/BA80 4B024/CA04 4B024/DA02 4B024/DA05 4B024/DA06 4B024/DA11 4B024/DA12 4B024 /EA04 4B024/GA11 4B024/HA12 4B063/QA18 4B063/QA19 4B063/QQ03 4B063/QQ08 4B063/QQ43 4B063/QR08 4B063/QR42 4B063/QR56 4B063/QS25 4B063/QS34 4B063/QX02 4B063/QX07 4B064 /AG26 4B064/CA10 4B064/CA20 4B064/CC24 4B064/DA01 4B064/DA13 4B065/AA01X 4B065/AA57X 4B065/AA72X 4B065/AA90X 4B065/AA93Y 4B065/AB01 4B065/BA02 4B065/CA24 4B065/CA44 4B065/CA46 4C084/AA02 4C084/AA06 4C084/AA07 4C084/AA13 4C084/AA17 4C084/BA01 4C084 /BA08 4C084/BA22 4C084/BA23 4C084/CA17 4C084/MA01 4C084/NA14 4C084/ZB212 4C084/ZB262 4C085/AA13 4C085/AA14 4C085/CC32 4C085/EE01 4C085/GG01 4C086/AA01 4C086/AA02 4C086 /AA03 4C086/AA04 4C086/EA16 4C086/MA01 4C086/MA04 4C086/NA14 4C086/ZB21 4C086/ZB26 4H045/AA10 4H045/AA11 4H045/AA20 4H045/AA30 4H045/BA10 4H045/CA40 4H045/DA75 4H045 /EA28 4H045/EA51 4H045/FA72 4H045/FA74		
优先权	60/215454 2000-06-30 US		
外部链接	Espacenet		

摘要(译)

本发明提供了编码ECM相关蛋白的cDNA。还提供了cDNA，片段，互补物，变体和编码的蛋白质或其部分，用于诊断和治疗疾病，特别是结肠癌和肺癌，以及该抗体的用途。本发明还提供了用于产生蛋白质和遗传转化模型系统的表达载体和宿主细胞。

	10		19		28		37		46		55						
G	CEA	TIG	AGA	GTC	TTC	CPC	CAA	ACA	TCT	TOC	GAT	TIG	TTC	CPT	TAA	CCC	ATC

	64		73		82		91		100		109						
G	ATC	TTC	GTC	GAA	ATC	AAT	TAC	AAA	CAT	TGC	CTT	ATG	TTC	GTT	TTC	TCG	AAC

	118		127		136		145		154		163						
A	TTC	GTC	GAA	TAT	TGG	ATC	TTC	AGT	TGG	AGG	ACA	ACA	ANT	GGG	CCT	GCA	ATC

	172		181		190		199		208		217						
G	ACT	TAT	TGC	AGT	TAA	AAA	CTT	GGT	TGG	AGA	ACA	TGC	CTC	CAC	AGT	CTA	TAA

	226		235		244		253		262		271						
T	GGT	GAT	GTT	CTC	TGC	AAC	AGC	CCT	CCA	TTT	TTC	AAA	GGA	AGT	ATA	CTC	AGT

	280		289		298		307		316		325						
A	CTA	AAG	AAG	GAA	TCT	ATT	TGC	CCT	ACT	CCA	CCA	GTG	TAT	GAA	GAA	CAT	GAG

	334		343		352		361		370		379						
T	CCT	TCA	GGA	TCA	TTA	CAT	CTG	GCA	GCA	ACA	TCT	TCA	ATA	AAC	GAT	AGT	CGC
