

(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2004-537971

(P2004-537971A)

(43) 公表日 平成16年12月24日(2004.12.24)

(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
C 0 7 K 1/02	C 0 7 K 1/02	2 G O 4 5
C 0 7 K 14/82	C 0 7 K 14/82	4 B O 2 4
C 0 7 K 16/32	C 0 7 K 16/32	4 B O 6 3
C 1 2 N 1/15	C 1 2 N 1/15	4 B O 6 4
		4 B O 6 5
	審査請求 未請求 予備審査請求 有	(全 132 頁) 最終頁に続く

(21) 出願番号 特願2002-559796 (P2002-559796)
 (86) (22) 出願日 平成13年12月13日 (2001.12.13)
 (85) 翻訳文提出日 平成15年6月16日 (2003.6.16)
 (86) 国際出願番号 PCT/US2001/048917
 (87) 国際公開番号 W02002/059311
 (87) 国際公開日 平成14年8月1日 (2002.8.1)
 (31) 優先権主張番号 09/740, 235
 (32) 優先日 平成12年12月18日 (2000.12.18)
 (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 ラセック、エイミー・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・フォーティーフ
 アースストリート 237

最終頁に続く

(54) 【発明の名称】 大腸癌マーカー

(57) 【要約】

本発明は大腸癌マーカーをコードするcDNAを提供する。
 本発明はまた、cDNAの利用のために、自身の断片、補体
 、及び変異体を提供し、コードされたタンパク質の利用
 のために、とりわけ大腸癌及びポリープのような結腸疾
 患の診断及び治療のために自身の一部及び抗体を提供す
 る。本発明はまた、タンパク質及び遺伝子組換モデルシ
 ステムの作成のための発現ベクターおよび宿主細胞をも
 提供する。

```

          9       18       27       36       45       54
5'  GCC TGC CGA GGT CCG AGC GAC CGA TGG AGA TGG CGG CTG CCG CTG AGT GAC GGA
          63       72       81       90       99
    CGG TGG AGG CCC AGA GCC CGG GCC TGA AGG GGG GGA CAA ACC TGG GTG CCC GCA
          117      126      135      144      153      162
    GGA GCC CCG CAG GGT GTC TTA CAA GTA TCA AGA ACT TAC TAT ATG TGG TTG AAT
          171      180      189      198      207      216
    AAA CAA TCA AGG TAA AGA GCA TCA AGT AAA AAC TTC TCC TTG TTA AGT ACT
          225      234      243      252      261      270
    TCA GAC ATT CCC CCA GTG GCT GAA GTG GCA TAT GAA TTA TGA AGT TGG ATC ATT
          279      288      297      306      315      324
    TGG AAT GAA TGT AAG AGA ATT GCC AAG GGC TCC TCC TCC AGA GAG GAA ACC
          333      342      351      360      369      378
    TCA TCC AGG GCC ATG AAG CCA CTT CCT CAC CAT CTG TGT GCT GCT TAA GCT AAT
          387      396      405      414      423      432
    GCT GCG GGA ACC ATG GTT CCT TGG GAG GAA TCA AGC TGA CTC TTG GCA TGA GAT
    
```

【特許請求の範囲】

【請求項 1】

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単離されたcDNA。

【請求項 2】

単離されたcDNAであって、

(a) SEQ ID NO:2の核酸配列、若しくはその補体と、

(b) SEQ ID NO:3乃至22から選択されたSEQ ID NO:2の断片、若しくはその補体と、

(c) SEQ ID NO:23乃至34から選択されたSEQ ID NO:2の変異体とから選択される単離されたcDNA。

【請求項 3】

前記請求項 1 のcDNA若しくはその補体を有する組成物。

【請求項 4】

前記請求項 1 のcDNAを含むベクター。

【請求項 5】

前記請求項 4 のベクターを有する宿主細胞。

【請求項 6】

タンパク質を生成するためにcDNAを用いる方法であって、

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

(b) 前記宿主細胞の培養から前記タンパク質を回収する過程とを含む方法。

【請求項 7】

サンプル中の核酸の発現を検出するのにcDNAを用いる方法であって、

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

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

【請求項 8】

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

【請求項 9】

前記組成物が基質に取り付けられることを特徴とする請求項 7 の方法。

【請求項 10】

前記cDNAが前記標準物質と比較される場合に差動的に発現され、直腸癌若しくは結腸ポリープの指標となることを特徴とする請求項 7 の方法。

【請求項 11】

複数の分子若しくは混合物をスクリーニングするためにcDNAを用いる方法であって、特異的な結合を可能とするための条件下で、前記請求項 1 のcDNAと複数の分子若しくは混合物とを組み合わせる過程と、

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

【請求項 12】

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

【請求項 13】

精製されたタンパク質若しくはその一部であって、

(a) SEQ ID NO:1のアミノ酸配列と、

(b) SEQ ID NO:1の抗原性エピトープと、

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

【請求項 14】

10

20

30

40

50

前記請求項 1 3 のタンパク質を含む組成物。

【請求項 1 5】

少なくとも一つのリガンドで同定するべく、複数の分子若しくは混合物をスクリーニングするためにタンパク質を用いる方法であって、

特異的な結合を可能とするための条件下で、前記請求項 1 3 のタンパク質と前記分子若しくは混合物とを組み合わせる過程と、

特異的な結合を検出し、それにより前記タンパク質に特異的に結合するリガンドを同定する過程とを含む方法。

【請求項 1 6】

前記分子若しくは混合物が、DNA分子、RNA分子、ペプチド核酸、ペプチド、タンパク質、mimetic、アゴニスト、アンタゴニスト、抗体、免疫グロブリン、阻害剤、及び薬剤から選択されることを特徴とする請求項 1 5 に記載の方法。

10

【請求項 1 7】

抗体を調整し精製するためにタンパク質を用いる方法であって、

(a) 抗体反応を誘発する条件下で、請求項 1 3 のタンパク質で動物を免疫する過程と、

(b) 動物の抗体を単離する過程と、

(c) 前記タンパク質を基質に取りつける過程と、

(d) 前記タンパク質に対する特異的結合が可能となるような条件下で、前記基質と単離した抗体とを接触させる過程と、

(e) 前記抗体を前記タンパク質より解離させ、それによって精製された抗体を獲得する過程とを含む方法。

20

【請求項 1 8】

前記請求項 1 7 の方法によって生成された抗体。

【請求項 1 9】

SEQ ID NO:1を有する単離されたタンパク質。

【請求項 2 0】

前記請求項 1 9 のタンパク質に特異的に結合された精製された抗体。

【請求項 2 1】

前記請求項 2 0 の抗体の特異性を有する多クローン性抗体を調整し精製するための方法であって、

30

(a) 抗体反応を誘発する条件下で、動物をSEQ ID NO:1のタンパク質で免疫する過程と、

(b) 動物の抗体を単離する過程と、

(c) 前記タンパク質を基質に取りつける過程と、

(d) 前記タンパク質に対する特異的結合が可能となるような条件下で、前記基質と単離した抗体とを接触させる過程と、

(e) 前記抗体を前記タンパク質より解離させ、それによって精製された多クローン性抗体を獲得する過程とを含む方法。

【請求項 2 2】

前記請求項 2 1 の方法によって生成された多クローン性抗体。

40

【請求項 2 3】

前期請求項 2 0 の抗体の前記特異性を有する単クローン性抗体を調整するための方法であって、

(a) 抗体反応を誘発する条件下でSEQ ID NO:1のタンパク質で動物を免疫する過程と、

(b) 前記動物より抗体生成細胞を単離する過程と、

(c) 培養内で前記抗体生成細胞と不死化細胞とを融合させ、ハイブリドーマ細胞を生成する単クローン性抗体を形成する過程と、

(d) 前記ハイブリドーマ細胞を培養する過程と、

(e) 培養より単クローン性抗体を単離する過程とを含む方法。

【請求項 2 4】

50

前記請求項 2 3 の方法によって生成された単クローン性抗体。

【請求項 2 5】

抗体を用いて、タンパク質を免疫精製するための方法であって、

(a) 基質に請求項 2 0 の抗体を付着させる過程と、

(b) 抗体：タンパク質複合体を形成可能とする条件下で、タンパク質を有するサンプルに対して前記抗体を露出させる過程と、

(c) 前記複合体より前記タンパク質を分離させる過程と、

(d) 前記精製タンパク質を回収する過程とを含む方法。

【請求項 2 6】

抗体を用いて、サンプル内でタンパク質の発現を検出するための方法であって、

(a) 前記抗体：タンパク質複合体の形成が可能な条件下で、前記請求項 2 0 の抗体とサンプルとを混合する過程と、

(b) 前記サンプル内で前記タンパク質の発現を示すことを特徴とする複合体形成物を検出する過程とを有する方法。

【請求項 2 7】

複合体形成物が基準物質と比較され、また、特に大腸癌及びポリープのような結腸疾患と診断されることを特徴とする請求項 2 に記載の方法。

【請求項 2 8】

請求項 2 0 の抗体及び標識部分を含む組成物。

【請求項 2 9】

請求項 2 0 の抗体及び薬剤を含む組成物。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本発明は、大腸癌マーカーをコードする哺乳動物 cDNA に関し、また、とりわけ大腸癌及びポリープのような結腸疾患に於ける診断及び治療で cDNA 及びコードされたタンパク質を用いることに関する。

【背景技術】

【0 0 0 2】

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

【0 0 0 3】

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, JD (1999) J Natl Cancer Institute 91:916-932を参照)。大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。また米国では大腸癌の診断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)。

【0 0 0 4】

大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大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。発現の変化が大腸癌の早期指標あるいは大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。

10

20

30

40

50

【0005】

例えば、DNAのメチル化の異常パターンがヒトの腫瘍において一貫して発生し、同時に広範な遺伝子の低メチル化及び増加したメチル化の限局性領域を含むことが良く知られている。とりわけ大腸癌では、これら変化が、大腸癌に先立つ前癌ポリープに於けるような腫瘍の進行の初期に発生することがわかっていた。実際、DNAの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、組織学的に大腸癌を有する患者よりのノーマル粘膜中、及び癌前に発生する良性のポリープ中において著しく増加する。この増加は結腸新生物が進行する間持続する (El-Deiryら (1991) Proc Natl Acad Sci USA 88:3470-3474)。増加したDNAのメチル化は、"CpG アイランド"と称されるゲノミックDNAのG+Cリッチ領域で発生する。それは、遺伝子周辺の"open"転写コンフォメーションのメンテナンスのために重要であり、また、これら領域のハイパーメチル化は、遺伝子の転写を鎮める"closed"コンフォメーションという結果になる。CpGアイランドのそのような異常メチル化による、分化遺伝子の沈黙化若しくは下方制御が、不死化細胞における分化を阻害しうることがこれまでに示唆されてきた (Antequeraら (1990) Cell 62:503-514)。

10

【0006】

家族性腺腫様ポリープ (FAP) は、大腸癌に先立ち、腺腫様ポリープ大腸菌 (APC) 遺伝子中で遺伝ミューテーションによって引き起こされる、微小常染色体優性症候群である。FAPは、平均44歳で癌へと進行する多発性結腸直腸アデノーマによって特徴づけられている。APC遺伝子は、APC-β-カテニン-Tcf (T-細胞因子) 経路の一部である。この経路の障害は、結果として、正しい順序での複製のロス、アポトーシス、及びポリープの成長を伴う結腸上皮性細胞の移動となる。一連のその他遺伝的变化はAPC-β-カテニン-Tcf経路の活性化に続き、通常結腸粘膜から転移性癌への転移を伴う。これらの変化は、K-Rasプロト腫瘍遺伝子の変異、メチル化パターン、及び腫瘍抑制遺伝子p53及びSmad4/DPC4の変異若しくはロスを含む。突然変異したAPC遺伝子の遺伝はまれにしか発生しないが、APC-β-カテニン-Tcf経路上のAPC及び結果的な効果のロス若しくは変異は、一般人の大部分の大腸癌にとって中心的であると考えられている。

20

【0007】

遺伝的非ポリープ性結腸直腸癌 (HNPCC) は、FAPより少量の良く定義されたフェノタイプを含む別の遺伝性常染色体優性症候群である。約2%の結腸直腸癌のケースの説明となるHNPCCは、癌の初期的な発現、及び、とりわけ子宮内膜、尿路、胃、及び胆管系に関連する別の癌の発生傾向によって識別される。HNPCCは、DNAミスマッチ修復 (MMR) 経路における一つ若しくは複数の遺伝子の変異に由来する。2つの人MMR遺伝子MSH2及びMLH1における変異は、現在までに同定された大部分のHNPCCファミリーに見られる。DNA MMR経路は、複製の間にDNAポリメラーゼの活性に由来するエラーを識別、修復する。さらには、MMR活性のロスが、アポトーシスを制御するBAX遺伝子や、セルの成長を制御するTGFβレセプターII遺伝子のロスのような、別の遺伝子の変異及び失欠の蓄積を通じて癌の進行に寄与する。DNA MMR異常を有する個体中のDNAに対する非回復性損傷のためのポテンシャルのために、癌腫への悪化が通常よりも早く進行する。

30

【0008】

潰瘍性大腸炎が、大腸癌に対するマイナーコントリビュータであり、患者は癌が成長するリスクが約20倍増加する。悪化は、組織学的に通常な組織においてさえも初期に発生しうるp53遺伝子における変異と関連を有する。中間体のポリープ状態のない潰瘍性大腸炎から異形成/癌腫に至る病気の進行は、結腸内容物に対する結腸粘液内増殖性細胞の発現から結果的に生じる高度な突然変異誘発性活動を示す。

40

【0009】

ほとんどすべての大腸癌は、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 (Issaら (1994) Nature Genetics 7:A536-540)。培養した大腸癌細胞に外来性ER遺伝子を導入することにより、著しく成長が抑制される。癌細胞侵襲の抑制は、ドメイン及びERのN-末端亜鉛フィンガー領域に結合したホルモン機能に依存する (Platetら (2000) Mol End

50

ocrinol 14:999-1009)。ホルモン結合によるERの活性化には特異的な標的遺伝子の転写が含まれ、また癌の侵襲性の減少に関連づけられても良い。ホルモンがない場合、亜鉛フィンガー領域によるタンパク質-タンパク質相互作用は、癌細胞の移動阻害に対しても貢献しうる。結腸上皮細胞中のERタンパク質の損失と癌の結果的発生との間の関連性は、立証されていない。

【0010】

FYVBフィンガータンパク質は、レセプタの情報伝達、小嚢性輸送、及びアクチン制御膜内外転位のような細胞プロセスで役割を果たす(Stemmark及びAasland (1999) J Cell Science 112:4175-4183)。FYVEドメインは、亜鉛フィンガーの一種であり、典型的には保存グリシン及びアルギニン残基を含む2つのZn²⁺カチオンに結合する8つの保存システインと、コンセンサス配列R(R/K)HHCRを有するベーシックモチーフとを有する。FYVEドメインは特異的な膜内にみられるホスホイノシチドに結合する。FYVEフィンガータンパク質中のその他のタンパク質の存在は、膜にその他の分子を有するタンパク質-タンパク質相互作用を媒介し、特定の細胞位置に対する小さなGTPaseのようなシグナル伝達分子の漸増をとともなっても良い。FYVEフィンガータンパク質中にヌルmutationを有するマウスでは、胎児は腹部の嚢形成形態発生過程の以上を示し、子宮内で死亡した(Komada及びSoriano (1999) Genes Dev 13:1475-1485)。卵黄嚢の腹部外側領域を有して成長した胎児は、2つの独立した両サイドの心臓管を有し、前腸を有しない。腸内に於ける細胞の分化及び増殖を制御する遺伝子中の変異は、結腸疾患との関連性を有するかもしれない。腸内でセル分化と増殖を制御する遺伝子における変異は、コロン病に関係するかもしれない(Doveら (1998) Phil Trans R Soc Lond B 353:915-923)。

10

20

【0011】

明確に、大腸癌、及び疾患の発生及び悪化に関連する多くの遺伝子変性が存在する。とりわけ、遺伝子の発現若しくは欠失の減少は、癌発達の初期的なインジケータを潜在的に提供し、疾患の悪化をモニターし、可能な治療ターゲットを提供するべく用いられても良い。所与のケースの大腸癌で影響を受ける特定遺伝子は疾患の分子進行に依存する。大腸癌に関連している追加遺伝子及び前癌状態の同定は、疾患の発生及び悪化に関するより信頼度の高いパターンを提供しうる。

【0012】

哺乳類の腸タンパク質をコードするcDNAの発見は、とりわけ大腸癌及びポリープのような結腸疾患の診断及び治療に有用な成分を提供することで、本技術分野の要求に応える。

30

【発明の開示】

【発明の効果】

【0013】

本発明は、特に大腸癌及びポリープのような結腸疾患の診断・治療に有用な大腸癌マーカー(CCM)をコードする哺乳動物cDNAの発見に基づいている。

【0014】

本発明は、SEQ ID NO:1のアミノ酸配列、SEQ ID NO:1のアミノ酸配列と少なくとも85%の同一性を有する変異体、SEQ ID NO:1の抗原性エピトープ、及びSEQ ID NO:1の生物学的活性化部位を含む群から選択されるほ乳類タンパク質またはその一部をコードする単離されたほ乳類タンパク質またはその断片を提供する。本発明はまた、SEQ ID NO:2の核酸配列、SEQ ID NO:3乃至22から選択されたSEQ ID NO:2の断片、及びSEQ ID NO:23-34から選択されたSEQ ID NO:2の核酸配列と少なくとも87%の同一性を有する変異体とを含む群から選択される単離された哺乳動物cDNAまたはそのその相補体を提供する。更に本発明は、CCMをコードするcDNAまたはその相補体を含む組成物、基板、及びプローブを提供する。更に本発明は、このようなcDNAを含むベクター、このようなベクターを含む宿主細胞、及びこのようなcDNAを用いてCCMを作製する方法を提供する。或る実施態様では、本発明は少なくとも一つのそれら断片を含む基板を提供する。第2の実施態様では、本発明は、検出法、スクリーニング法、及び精製法に用いることが可能な断片を含むプローブを提供する。更なる実施態様では、プローブは一本鎖相補RNAまたはDNA分子である。

40

50

【0015】

本発明はまた、cDNAを用いて、サンプルの核酸の発現変動を検出する方法であって、前記プローブと前記核酸とをハイブリダイズさせてハイブリダイゼーション複合体を形成し、そのハイブリダイゼーション複合体を標準と比較し、その比較によりサンプルにおける前記cDNAの発現変動が分かる方法を提供する。或る実施態様では、この検出方法は、ハイブリダイゼーションの前に前記核酸を増幅するステップを含む。別の実施態様では、この検出方法はcDNAの異なった発現が、特に大腸癌及びポリープのような結腸疾患の診断に用いられることを示している。また別の側面では、cDNA若しくは断片、若しくは補体はエレメント若しくはアレイを含んでも良い。

【0016】

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

【0017】

本発明はまた、SEQ ID NO:のアミノ酸配列と、SEQ ID NO:1のアミノ酸配列と85%の同一性を有する変異体と、SEQ ID NO:1の抗原エピートープと、SEQ ID NO:1の生物学的に活性部分とからなる群から選択される精製された哺乳類タンパク質またはその一部を提供する。本発明はまた、前記精製されたタンパク質、若しくは医薬用担体に関連する前記タンパク質の部分を含む組成物を提供する。本発明は更に、とりわけ大腸癌及びポリープのような結腸疾患を有する被験者を治療するためにCCMを用い、そのような治療を必要とする患者に生成したタンパク質を含む合成物を投与する方法を提供する。本発明は更に、タンパク質を用いてライブラリ即ち複数の分子または化合物をスクリーニングして少なくとも1つのリガンドを特定する方法であって、特異的な結合が許容される条件下で、そのタンパク質を前記分子または化合物と結合させるステップと、特異的な結合を検出して、前記タンパク質と特異的に結合するリガンドを特定するステップとを含む方法を提供する。ある実施態様においては、分子または化合物はDNA分子、RNA分子、ペプチド核酸、ペプチド、タンパク質、擬態、アゴニスト、アンタゴニスト、抗体、免疫グロブリン、インヒビター、及び薬剤から選択される。別の実施態様では、とりわけ大腸癌、及びポリープのような結腸疾患の患者を治療するために前記リガンドを使用する。

【0018】

本発明は、哺乳動物タンパク質を用いて、被検サンプルをそのタンパク質に特異的に結合する抗体に対してスクリーニングする方法であって、前記被検サンプルから抗体を単離するステップと、特異的な結合が許容される条件下で、この単離した抗体を前記タンパク質とを結合させるステップと、結合した前記タンパク質から抗体を解離するステップと、抗体の存在或いは存在量からとりわけ大腸癌、及びポリープのような結腸疾患の患者を診断できる抗体の量を既知の標準値と比較するステップとを含むスクリーニング法を提供する。

【0019】

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

【0020】

本発明は、とりわけ大腸癌、及びポリープのような結腸疾患において発現するタンパク質に特異的に結合する精製された抗体を提供する。本発明はまた、とりわけ大腸癌、及びポ

10

20

30

40

50

リープのような結腸疾患を診断する方法であって、抗体を混合して結合抗体量を既知の指標と比較するステップを含み、それによって、とりわけ大腸癌、及びポリープのような結腸疾患の存在を証明する診断方法を提供する。本発明は更に、抗体を用いてとりわけとりわけ大腸癌、及びポリープのような結腸疾患を治療する方法であって、精製された抗体を含む医薬組成物をそのような治療が必要な患者に投与することを含む治療方法を提供する。

【0021】

本発明はまた、哺乳動物のゲノムDNAの中にマーカー遺伝子を挿入して、内在性ポリヌクレオチドの発現を阻害する方法を提供する。本発明はまた、cDNAを用いて哺乳動物モデル系を作製する方法であって、SEQ ID NO:2-34から選択されるcDNAを含むベクターを作製するステップと、そのベクターで胚性幹細胞を形質転換するステップと、形質転換された胚性幹細胞を選択するステップと、この形質転換された胚性幹細胞を哺乳動物胚盤胞の中に微量注入して、キメラ胚盤胞を形成するステップと、偽妊娠メスにキメラ胚盤胞を導入し、このメスが、生殖細胞系の中にcDNAを含むキメラ子孫を出産し、そのキメラ哺乳動物を交配して同型接合哺乳動物モデル系を作製するステップとを含む哺乳動物モデル系作製方法を提供する。

10

【発明を実施するための最良の形態】

【0022】

(本発明の記載について)

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

20

【0023】

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

30

【0024】

(定義)

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

【0025】

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

40

【0026】

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

【0027】

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

【0028】

50

「タンパク質をコードするcDNA」という語は、当分野で周知の分析により同定された保存された領域、モチーフ或いはドメインをコードする配列と密接にアラインメントする核酸配列を指す。これらの分析には、保存された領域内における同一性を特定するBLAST (Basic Local Alignment Search Tool)が含まれる(Altschul (1993) J Mol Evol 36: 290-300、Altschul ら(1990) J Mol Biol 215:403-410)。

【0029】

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

10

【0030】

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

【0031】

「障害」とは、特に大腸癌及びポリープである結腸疾患のような、cDNA及びCCMが異なって発現されている場合の症状、疾患、または症候群を指す。

【0032】

「断片」とは長さが約200から約700塩基の連続したヌクレオチド鎖を指す。断片は、関連する核酸分子を同定するためにPCRまたはハイブリダイゼーション技術に、或いはリガンドをスクリーニングするための結合アッセイで使用され得る。核酸とそれらのリガンドはこのような方法で同定され、複製、転写、または翻訳を調節する治療として有用である。

20

【0033】

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

30

【0034】

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

【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や増幅技術において、または複製、転写、翻訳の調節において使用され得る。概ね等しい用語としては、アンプライマー、プライマー、及びオリゴマーが挙げられる。

40

【0036】

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

【0037】

タンパク質の「翻訳後修飾」には、脂質化、グリコシル化、リン酸化、アセチル化、ラセミ化、及び蛋白分解性切断等が含まれ得る。これらのプロセスは、合成或いは生化学的に生じ得る。生化学的修飾は、細胞の位置、細胞型、pH、及び酵素環境などによって異なる。

【0038】

「プローブ」とは、サンプル中の少なくとも1つの核酸にハイブリダイズするcDNAを指す

50

。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*法、ドットプロット法、及びアレイなどを含むハイブリダイゼーション反応またはスクリーニングアッセイで使用するためにレポーター分子で標識化することができる。

【0039】

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

10

【0040】

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

【0041】

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

20

【0042】

「特異的な結合」とは、構造、特に分子側鎖に依存する2つの分子間での特殊な正確な相互作用を指す。例えば、調節タンパク質のDNA分子の主溝への挿入、2つの一本鎖核酸間の骨格に沿った水素結合、またはタンパク質のエピトープとアゴニスト、アンタゴニストまたは抗体との間の結合がある。

【0043】

配列に適用される「類似性」とは、Smith-Waterman アルゴリズム(Smith及びWaterman (1981) J Mol Biol 147:195-197)或いはBLAST2 (Altschul ら (1997) Nucleic Acids Res 25:3389-3402)などの標準化されたアルゴリズムを用いてアラインメントされた少なくとも2つの配列間で一致する分子或いは残基の定量化(通常は%)を指す。BLAST2は、アラインメントを最適化するために配列の1つでギャップを挿入するまた2つの配列をより有意に比較できる標準化された再現性のある方法で使用され得る。

30

【0044】

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

【0045】

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

40

【0046】

(発明)

本発明は、大腸癌マーカーをコードする哺乳動物cDNAの発見と、結腸疾患の性質決定、診断、治療における、cDNAまたはその断片、およびタンパク質またはその部分をそのまま使用するか組成物として使用することに基づくものである。

50

【 0 0 4 7 】

本発明のCCMをコードする核酸は、アミノ酸配列のためのコンピュータサーチを用い、哺乳動物上皮性細胞cDNAライブラリ (EPIMNON05) よりインサイトクローン5098390内で最初に同定された。コンセンサス配列、SEQ ID NO:2は、追従するオーバーラッピング及び/又は拡張された核酸配列 (SEQ ID NOs:3-22) : インサイトクローン5098390F6 (EPDvINON05)、8022266J1、73940661H7、6309461HI (NERDTDN03)、7436032H-1、7676140H29、6923049H1 (PLACFER06)、8216794J1、6995895HI (BRAQTDR02)、6889439H1 (BRAITDR03)、70405975D1 (SG00000195)、8117068H1、747646H1 (BRAITUT01)、7984383H1、8073132J1、6744627H (BRAFN0T02)、6749637H1 (BRAXNOT03)、5073165F8 (COLCTUT03)、6441214H1 (BRAENOT02)、2436362H1 (BRAVUNT02)、GenBank EST g921283 (SEQ ID NO:35)及びゲノミック配列GNFL.g8572 864 000002_002.edit (SEQ ID NO:36)に由来した。表1は、マイクロアレイ解析で測定されるような健康な結腸組織に対する大腸癌及びポリープ中のCCMの特異発現を示すものである。ある患者の少なくとも二倍であり、大多数の患者の少なくとも2倍であれば、特異発現 (DE) は顕著であると考えられた。CCMは、大腸癌を有する患者より抽出した組織内で、同一患者から採取した顕微鏡学的に健康な組織と比較して軽減された発現を示す (Dn3753、Dn3649、及びDn3647)。CCMは、結腸腺腫を有する患者より抽出した組織内で、同一患者から採取した顕微鏡学的に健康な組織と比較して軽減された発現を示す (Dn3583)。CCMは、ポリープを有する患者より抽出した組織内で、同一患者から採取した顕微鏡学的に健康な組織と比較して軽減された発現を示す (Dn3753)。故に、cDNAは結腸疾患、とりわけ大腸癌及びポリープのような大腸癌の診断アッセイにおいて有用である。約1から50のヌクレオチドのcDNAの断片もまた診断アッセイに有用である。

【 0 0 4 8 】

ある実施例では、本発明はSEQ ID NO: 1のアミノ酸配列を含むポリヌクレオチドを包含する。CCMは、長さが784アミノ酸である。また、CCMは、N465に存在する1つのN-グリコシレーション部位、S230に存在する一つの潜在的環式AMP-若しくは環式GMP-依存タンパク質キナーゼリン酸化部位、S3、S31、S38、T132、T137、S239、T484、S635、S691、及びS726に存在する10の潜在的カゼインキナーゼIIリン酸化部位、S141、S171、S226、S333、S407、S429、T501、S508、及びS654に存在する8つの潜在的プロテインキナーゼCリン酸化部位を有する。MOTIFS解析は、H37を介したC16よりのCCM領域がC2H2型亜鉛フィンガーモチーフに類似することを示す。BLOCK解析は、M191を介したC182よりのCCM領域がC3HC4亜鉛フィンガー (RINGフィンガー) モチーフに類似することを示す。Pfam解析は、K261を介したV152よりのCCM領域がFYVEフィンガードメイン ($E\text{-value} = 3.8 \times 10^{-15}$) に類似することを示す。特に、CCMはC163、C166、C179、C182、C187、C190、C252、C255に8つの保存システイン残基を有し、残基R179を通して残基R175にRRHHCRコンセンサス配列を有し、FYVEフィンガータンパク質を代表するG183における保存グリシンを有する。有用な抗原性エピソードは、N156からM186、N488からL525、及びA680からA730に広がっている。また、CCMの生物学的に活性な部位は、C16からH37、及びV152からK261に広がっている。CCMへ特異的に結合する抗体は、とりわけ大腸癌及びポリープのような結腸疾患に対するアッセイにおいて有用である。

【 0 0 4 9 】

本発明は、CCM変異体も包含する。好ましいCCM変異体は、CCMアミノ酸配列に対して少なくとも約80%、若しくは約90%、さらには約95%のアミノ酸配列同一性を有し、また、少なくとも一つのCCMの機能的若しくは構造的特徴を含むものである。

【 0 0 5 0 】

CCMをコードするcDNAの哺乳動物変異体は、デフォルトのパラメータ及びZOOSEQデータベース (Incyte Genomics) と共にBLAST2を用いて同定された。これら好適な変異体は、以下の表に示すようにヒトCCMに対して約87%から約95%の同一性を有する。第1列は、ヒトcDNAのためのSEQ ID (SEQ IDH) を示しており、第2列は、変異体cDNAのためのSEQ ID (SEQ IDvar) を示しており、第3列は変異体cDNAのためのクローン番号 (Clonevar) を示しており、第4列は、ライブラリの名前を示しており、第5列は、ヒトcDNA (同じケースでヒ

トcDNAの異なった領域を含む、変異体cDNAの異領域配列を有する)に対する変異体の配列を示しており、そして第6列はヒトcDNAに対する同一性のパーセンテージを示している。

【0051】

SEQ ID _h	SEQ ID _{var}	クローン _{var}	ライブラリ名	N _h 配列	同一性
2	23	702073518T2	RAKITXT09	949-1543	93%
2	24	702549617T1	RACONON05	994-1544	92%
2	25	702448734T1	RASPNON05	1093-1544	93%
2	26	702081830H1	RABRTXT01	958-1376	94%
2	27	702053775T1	RAKITXT08	1151-1442	91%
2	28	700883909H1	RAVANOT01	721-975	90%
2	29	700883983H1	RAVANOT01	721-970	90%
2	30	701092129H1	RALUNOT02	1765-1967	88%
2	31	702137890H2	RABRFET07	397-613	87%
2	32	701741076T1	MNBCNON01	6021-6451	95%
2	33	702769948H2	CNLINOT01	1377-1833	88%
2	34	702245053H1	CNLUNOT01	3494-3542	93%

10

【0052】

これらcDNAは、ヒトの疾患を形成しそのような疾患の潜在的治療をテストする遺伝子組み換え細胞株若しくは生物体を生成するのにとりわけ有用である。

【0053】

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

20

【0054】

cDNA及びその断片自身 (SEQ ID NOs: 2-34) は、SEQ ID NO: 1およびサンプル内の関連分子間で同定及び識別可能する合成、増幅、及びスクリーニング技術において用いられても良い。ほ乳類のcDNAは、遺伝子組み換え細胞株若しくは生物を生成するのに用いられ得る。それらは、ヒトの結腸疾患とりわけ大腸癌及びポリープのためのモデルシステムであり、潜在的な治療の毒性及び効果がテストされても良い。毒物学の学習、臨床試験、及び被験者/患者の処置のプロフィールは、cDNA、タンパク質、抗体、分子および化合物を用いて実行及びモニターされても良い。前記分子及び化合物は、本発明の分子及びcDNAやタンパク質を用いて識別されたものである。

30

【0055】

(本発明の特徴及び使用)

cDNAライブラリ

ここに開示する特定の実施例では、当分野で周知の方法を用いて哺乳動物の細胞及び組織からmRNAを単離し、これを用いてcDNAライブラリを作製する。上記したインサイト社クローンは、哺乳動物cDNAライブラリから単離された。本発明の代表的な3つのライブラリの作製方法を以降に記載する実施例に示す。コンセンサス配列は、Phrap (P. Green, University of Washington, Seattle WA) 及びAUTOASSEMBLERアプリケーション (Applied Biosystems, Foster City CA) などのコンピュータプログラムを用いて、インサイト社クローンを含む断片、伸長、及び/またはショットガン配列から化学的かつ/または電子的に構築した。クローン、伸長、及び/またはショットガン配列は、クラスタ及び/又はマスタークラスタへと電子的に集まる。

40

【0056】

シークエンシング

核酸をシークエンシングする方法は当分野で周知であり、そのような方法を用いて本発明の任意の実施例を実施することができる。これらの方法は、DNAポリメラーゼIであるクレ

50

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

10

【0057】

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

20

【0058】

核酸配列の伸長

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

30

【0059】

ハイブリダイゼーション

cDNA及びその断片は、様々な目的のための様々なハイブリダイゼーション技術に用いることができる。プローブは、5'調節領域や非保存領域 (すなわち、タンパク質の保存された触媒ドメインをコードする5'若しくは3'のヌクレオチド) 等のユニークな領域から作製可能であり、CCM、対立遺伝子の変種、若しくは関連分子をコードする天然の分子を同定するためのプロトコルに用いることができる。通常は一本鎖であるDNA若しくはRNAからなるこのプローブは、任意のこの核酸配列、配列番号2-34と少なくとも50%の配列同一性を有することが望ましい。ハイブリダイゼーションプローブは、レポータ分子の存在下でのPCR増幅、オリゴ標識化、ニックトランスレーション法、または末端標識化を利用して作製することができる。このcDNAまたはその断片を含むベクターを用いて、RNAポリメラーゼ及び標識したヌクレオチドを加えて *in vitro* でmRNAプローブを作製することができる。これらの方法はAPBが販売するキットを用いて行うことができる。

40

【0060】

ハイブリダイゼーションのストリンジェンシー (厳密性) は、プローブのGC含量、塩濃度、及び温度によって決まる。特に、塩濃度を下げる、またはハイブリダイゼーションの温

50

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

10

【0061】

当分野で周知の方法でマイクロアレイを準備して分析することができる。オリゴヌクレオチドをマイクロアレイのプロブや標的として用いることができる。マイクロアレイを用いて、同時に極めて多数の遺伝子の発現レベルをモニタリングし、遺伝子変異体、突然変異及びSNP（一塩基多型）を同定することができる。このようなデータを用いて、遺伝子機能の解明や、症状及び疾患、または障害における遺伝子原理の解明や、症状及び疾患、障害の診断または治療、治療薬の開発、並びにこれらの治療薬の活性のモニタリングが可能である（例えば、Brennan他(1995) USPN 5,474,796; Schena他(1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler他(1995) PCT出願 W095/251116; Shalon他(1995) PCT出願 W095/35505; Heller他(1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller他(1997) USPN 5,605,662を参照）。

20

【0062】

ハイブリダイゼーションプロブはまた、天然のゲノム配列のマッピングに有用である。このプロブは、（1）特定の染色体、（2）染色体の特定の領域、（3）ヒト人工染色体（HACs）や酵母人工染色体（YACs）、細菌人工染色体（BACs）、細菌P1作製物、若しくは単一クロモソームcDNAライブラリのような人工染色体作製物、にハイブリダイズすること可能である。

30

【0063】

発現

CCMをコードする複数のcDNAの何れか一つをベクターにクローニングして、このタンパク質若しくはその一部を宿主細胞で発現させることができる。この核酸配列を、DNAシャッフリング（Stemmer and Cramer (1996) USPN5,830,721）や部位特異的変異誘発などの方法によって、新規の制限部位を作り出したり、グリコシル化パターンを変えたり、優先コドンを変えて特定の宿主における発現を増大させたり、スプライスバリエーションを作り出したり、半減期を延長する等の操作が可能である。この発現ベクターは、特定の宿主における各要素の効率に基づいて選択された様々なサンプルに由来する転写及び翻訳調節エレメント（プロモーター及びエンハンサー、特定の開始シグナル、ポリアデニル化3'配列）を含み得る。in vitro組み換えDNA技術、合成技術及び/またはin vivo遺伝子組み換え技術を組み合わせて、このベクターにcDNAと調節エレメントをつなぐことができる。このような技術は、当分野で周知であり、Sambrook(前出、ch. 4, 8, 16 and 17)に記載されている。

40

【0064】

様々な宿主系を発現ベクターで形質転換することができる。以下に限定するものではないが、これらの中には組み換えバクテリオファージやプラスミド、またはコスミドDNA発現ベクターで形質転換された細菌と、酵母発現ベクターで形質転換された酵母と、バキュロ

50

ウイルス発現ベクターで形質転換された昆虫細胞系と、ウイルスエレメント及び/または細菌エレメントを含む発現ベクターで形質転換された植物細胞系や動物細胞系が含まれる (Ausubel前出、unit 16)。例えば、アデノウイルス転写/翻訳複合体を哺乳動物細胞に用いることができる。配列をウイルスのゲノムのE1若しくはE3領域に結合させた後、この感染ウイルスを用いて形質転換させ、宿主細胞でタンパク質を発現させることができる。また、ラウス肉腫ウイルスエンハンサーやSV40、またはEBV系のベクターを用いてタンパク質を高発現させることができる。

【0065】

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

10

【0066】

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

20

【0067】

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

30

【0068】

細胞培地からのタンパク質の回収

精製を容易にするために、ベクターに導入する異種部分は、グルタチオンSトランスフェラーゼ (GST)、6-His、FLAG、MYC等を含む。GST及びCBP、6-Hisはそれぞれ、グルタチオン及びカルモジュリン、金属キレート樹脂が結合した市販のアフィニティマトリックスを用いて精製される。FLAG及びMYCは、市販のモノクローナル抗体及びポリクローナル抗体を用いて精製される。精製に続く分離を容易にするために、タンパク質の分割部位をコードする配列は、タンパク質と異異形部分との間のベクターの一部であって良い。組み換えタンパク質の発現及び精製の方法はAusubel(前出、unit 16)に記載され市販されている。

40

【0069】

ペプチドの化学合成

タンパク質若しくはその一部は、組み換え方法以外の当分野で周知の化学的方法によって合成することもできる。固相技術を用いるペプチド合成は、バッチ式或いは連続的なフロープロセスによって行うことができる。連続的なフロープロセスでは、アミノ保護及び

50

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

10

【0070】

抗体の準備及びスクリーニング

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

20

【0071】

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

30

【0072】

別法では、当分野で周知の方法を用いる上記した一本鎖抗体を生産する技術で、エピトープ特異的一本鎖抗体を生産する。タンパク質のエピトープに対して特異的に結合する部位を含む抗体断片を生産することが可能である。限定するものではないが、このような断片には、例えば、抗体分子のペプシン消化によって作製されたF(ab')₂断片及びこのF(ab')₂断片のジスルフィド架橋を減少させて作製したFab断片が含まれる。別法では、Fab発現ライブラリを作製して、目的の特異性を有するモノクローナルFab断片の高速かつ容易に同定できるようにする(例えば、Huse他(1989) *Science* 246:1275-1281を参照)。

40

【0073】

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

50

ッセイが好ましいが、競合的結合によるアッセイを用いることもできる (Pound (1998) Immunochemical Protocols. Humana Press, Totowa NJ)。

【0074】

アッセイのための分子の標識化

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

10

【0075】

(診断)

本cDNA、断片、オリゴヌクレオチド、相補的なRNA及びDNA分子、PNAを用いて、遺伝子発現の変化やmRNAの過剰な発現の不在/存在を検出及び定量、または治療期間中のmRNAレベルのモニタリングを行うことができる。IPと特異的に結合する同様の抗体は、タンパク質を定量化するのに用いられてもよい。発現変動に関連する疾患には、大腸癌、クーロン病、及び潰瘍性大腸炎が含まれる。診断アッセイにハイブリダイゼーションまたは増幅技術を用いて、遺伝子発現の変化を検出するべく、患者からの生体サンプルの遺伝子発現レベルを標準的なサンプルの値と比較する。質的または量的なこのような比較法は当分野で周知である。

20

【0076】

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

【0077】

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

30

【0078】

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

40

【0079】

免疫学的方法

特異的なポリクローナル抗体若しくはモノクローナル抗体の何れかを用いるタンパク質の検出及び定量は当分野で周知である。このような技術には、ELISA (酵素結合免疫吸着検定法) 及びラジオイムノアッセイ (RIA)、蛍光活性化セルソーター法 (FACS) が含まれ

50

る。2つの非干渉エピトープに反応するモノクローナル抗体を用いる2部位モノクローナル系イムノアッセイが好ましいが、競合的結合アッセイを用いることもできる(例えば、Coligan 他 (1997) Current Protocols in Immunology, Wiley-Interscience, New York NY; および Pound 前出)。

【0080】

(治療)

CCMの領域 (SEQ ID NO: 1) は、FYVEフィンガードメイン含有タンパク質に対して化学的・構造的な同一性を有する。加えて、CCMの異なる発現は表1に記載されているようにとりわけ大腸癌及びポリープと言った結腸疾患に高く関連している。CCMは、とりわけ大腸癌及びポリープのような結腸疾患に於いて際だった役割を有する。

10

【0081】

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

【0082】

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

20

【0083】

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

30

【0084】

核酸を用いる遺伝子発現の調節

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

40

【0085】

また、酵素活性をもつRNA分子であるリボザイムを用いて、RNAの特異的な切断を触媒してもよい。リボザイム作用のメカニズムは、まずリボザイム分子と相補的な標的RNAとの配列特異的なハイブリダイゼーションが起こり、次にGUAおよびGUU、GUCなどの部位においてヌクレオチド鎖が切断される。このような部位が一旦同定されたら、オリゴヌクレオチドを機能不全にしうる二次構造特性について、同じ配列を有するオリゴヌクレオチドを評価することができる。また、候補標的としての適合性は、RNA分解酵素保護アッセイを用

50

いて、相補的なオリゴヌクレオチドとのハイブリダイゼーションを検査して評価することができる。

【0086】

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

10

【0087】

スクリーニング及び精製のアッセイ

本CCMをコードするcDNAを用いて、分子若しくは混合物のライブラリをスクリーニングして特異的な結合親和性を調べることが可能である。このライブラリは、生物系においてcDNAの活性、複製、転写、または翻訳を調節するアプタマー、DNA分子、RNA分子、PNA、ペプチド、転写因子などのタンパク質、エンハンサー、リプレッサー、およびその他のリガンドを含み得る。このアッセイは、特異的な結合が許容される条件下で、cDNAまたはその断片を分子のライブラリと結合させるステップと、特異的な結合を検出して、一本鎖または適切ならば二本鎖分子と特異的に結合する少なくとも1つの分子を同定するステップとを含む。

20

【0088】

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

30

【0089】

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

【0090】

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適切なカオトロピック剤を使用することが含まれる。

40

【0091】

好適な実施例では、CCMまたはその部分を多様なスクリーニングアッセイの任意の複数の分子または化合物をスクリーニングするために用い得る。そのようなスクリーニングで用いられたタンパク質の部分は、溶液中で遊離しているか、非生物的物质または生物的基質に固定させる(例えば細胞表面上に保持される)、あるいは細胞内に位置することになる

50

う。例えば1つの方法では、組換え核酸で安定的に形質転換された、生きてまたは固定された原核宿主細胞で、細胞表面にペプチドを発現し局在させる細胞をスクリーニングアッセイで使用することが可能である。細胞を複数のリガンドあるいはリガンドのライブラリに対してスクリーニングする。そして発現したタンパク質とリガンドの間で結合の特異性または複合体の形成を測定し得る。タンパク質と分子間の特異結合を測定してもよい。スクリーニングされたライブラリの種類に応じて、アッセイをDNA分子、RNA分子、ペプチド核酸、ペプチド、タンパク質、擬態、アゴニスト、アンタゴニスト、抗体、免疫グロブリン、インヒビター、薬剤及び他のリガンドを同定するために用いる可能性があり、そしてタンパク質を特異結合する。

【0092】

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

【0093】

薬理学

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

【0094】

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

【0095】

（モデル系）

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

【0096】

毒性研究

毒性研究は、生物系上での薬剤影響の研究である。多くの毒性研究は、ラットまたはマウスで実施される。生理学、行動、恒常性プロセスにおける定性または定量変更そしてラットまたはマウスの死亡率の観察は、毒性プロフィールを生み出し、薬剤への暴露の後にヒ

10

20

30

40

50

トの健康に及ぼす潜在的な因果関係を評価するために使われる。

【0097】

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。

【0098】

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

【0099】

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

【0100】

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

【0101】

遺伝子組換え動物モデル

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

【0102】

胚性幹細胞

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

【0103】

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。

【0104】

ノックアウト分析

10

20

30

40

50

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

【0105】

ノックイン分析

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

【0106】

ヒト以外の霊長類モデル

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

【0107】

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

【実施例】

【0108】

下記の実施例は、目的の本発明を説明するために提供するが、これは本発明を限定する目的ではない。実施例の目的のために、ヒト脳後壁皮質(BRAENOT02)、及び標準化された乳房細胞(EPIMNON05)細胞の準備について記載する。

【0109】

1 cDNAライブラリの作製

脳後壁皮質ライブラリ

BRAENOT02 cDNAライブラリは、35歳の白人男性の脳より摘出した脳後壁皮質組織より作成された。POLYTRONホモジナイザー(Brinkmann Instruments, Westbury NJ)を使用して、凍結組織をTRIZOL試薬(0.8gの組織/12 mlのTRIZOLの割合; Life Technologies)でホモジナイズして溶解させる。氷上での短いインキュベーションの後に、クロロホルムを加え(1:5 v/v)、混合液を遠心分離機にかけて相に分離した。上部の水相を新しい試験管に移して、イソプロパノールを加えてRNAを沈殿させた。RNAをRNアーゼのない水で再懸濁し、DNアーゼで処理した。RNAを酸性フェノール-クロロホルムで再抽出し、酢酸ナトリウムとエタノールで再沈殿した。OLIGOTEX精製キット(QIAGEN, Chatsworth CA)を用いて、ポリ(A+) RNAを単離した。

【0110】

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

【0111】

標準化乳房上皮性細胞ライブラリ

実施例の目的のために、乳房上皮性細胞ライブラリ(EPIMN05)の標準化について記載する。E. coli ストレイン DH12S コンピテント細胞(Life Technologies)内のEPIMN05プラスミドライブラリ、約 3.28×10^6 個の独立クローンが、エレクトロポレーションによる形質転換に続くカルベニシリン(25 mg/l)及びメチシリン(1mg/ml)の選択条件下で液体培養中に生長する。ライブラリ内の存在レベルに従い過剰cDNAの複製数を減じるために、cDNAライブラリは、Soaresら(1994, Proc Natl Acad Sci 91:9228-9232)の手法に従ってシングルラウンド(single round)で標準化され、追従する変性をともなった。プライマーの拡張反応(primer extension reaction)におけるプライマー-テンプレート比は、2:1から10:1に上昇した。反応におけるdNTP濃度は各々のdNTPのために $150 \mu\text{M}$ まで減じられ、より長い(400から1000nt)のプライマー拡張生成物の生成を可能とする。リアニリング(reannealing)ハイブリダイゼーションは、13時間から48時間に延長された。標準化されたライブラリのシングルのスタンダードDNA循環(circle)は、ヒドロキシアパタイトクロマトグラフ法によって精製され、ランダムな初回抗原刺激によって部分的に二重鎖に変換され、また、DH12S適格細胞(Life Technologies)へと電気穿孔された。

【0112】

2 cDNAライブラリの標準化

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

【0113】

バクテリア(pSPORT 1 RI)により産出した仲介プラスミドはEcoRIでは消化されず、Hind III(New England Biolabs)で消化され、張出部(overhang)末端は再びクレノウとdNTPで占めた。リンカー配列をリン酸化して、5'平滑末端上に連結反応させ、EcoRIで消化してから、自己連結反応させた。JM109宿主細胞への形質転換後に、プラスミドを単離してEcoRIとの優先的消化性を検査したが、Hind IIIとの優先的消化性を検査しなかった。この基準にかなう単一のコロニーを、pINCYプラスミドと名付けた。

【0114】

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

【0115】

3 cDNAクローンの単離とシーケンシング

プラスミドDNAを細胞から放出して、MINIPREPキット(Edge Biosystems, Gaithersburg MD)またはREAL PREP 96 プラスミドキット(Qiagen)のいずれかを用いて精製した。このキットには、960精製のため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25 mg/lと0.4%のグ

10

20

30

40

50

リセロールの入った、殺菌したTERRIFIC BROTH(BD Biosciences, Sparks MD)1 mlで培養した。2) 接種後、細胞を19時間培養し、次いで0.3 mlの溶解バッファで溶解した。3) イソプロパノールによる沈殿の後、プラスミドDNAペレットを0.1 mlの蒸留水で再懸濁した。プロトコルの最終ステップ後に、サンプルを96-ウェルブロックに移して、4℃で保管した。

【0116】

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

10

【0117】

4 cDNA配列の伸長

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

20

【0118】

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。

【0119】

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、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℃で15秒間、ステップ 3:60℃で1分間、ステップ 4:68℃で2分間、ステップ 5:ステップ 2、3及び4を20回繰り返す。ステップ 6:68℃で5分間、ステップ 7:4℃で保管。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℃で15秒間、ステップ 3:57℃で1分間、ステップ 4:68℃で2分間、ステップ 5:ステップ 2、3及び4を20回繰り返す。ステップ 6:68℃で5分間、ステップ 7:4℃で保管。

30

40

【0120】

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

【0121】

50

伸長させたクローンは、脱塩及び濃縮して384穴プレートに移し、CviJIコレラウイルスエンドヌクレアーゼ (Molecular Biology Research, Madison WI) を用いて消化し、pUC 18ベクター (APB) への再連結反応前に音波処理またはせん断した。ショットガン・シーケンシングのために、消化したヌクレオチド配列を低濃度 (0.6 ~ 0.8%) のアガロースゲル上で分離し、断片を切除し、寒天をAGARACE酵素 (Promega) で消化した。伸長させたクローンをT4DNA (New England Biolabs) を用いてpUC 18ベクター (APB) に再連結し、Pfu DNAポリメラーゼ (Stratagene) で処理して制限部位の張出部 (overhang) を満たし、大腸菌細胞に形質移入した。形質移入した細胞を選択して抗生物質を含む培地に移し、それぞれのコロニーを切りとってLB/2Xカルベニシリン培養液の384ウェルプレートに37 で一晩培養した。

10

【0122】

細胞を溶解して、Taq DNAポリメラーゼ (APB) 及びPfu DNAポリメラーゼ (Stratagene) を用いて以下の手順でDNAをPCR増幅した。ステップ 1:94 で3分間、ステップ 2:94Cで15秒間、ステップ 3:60Cで1分間、ステップ 4:72Cで2分間、ステップ 5:ステップ 2、3及び4を29回繰り返す。ステップ 6:72Cで5分間、ステップ 7:4 で保管。上記したようにPICOGREEN量的試薬 (Molecular Probes) でDNAを定量化した。DNAの回収率が低いサンプルは、上記と同一の条件を用いて再増幅した。サンプルは20%ジメチルスルホキシド (DMSO; 1:2, v/v) で希釈し、DYENAMIC エネルギートランスファー シーケンシングプライマー、及びDYENAMIC DIRECTサイクルシーケンシングキット (APB) またはABI PRISM BIGDYE ターミネーターサイクル シーケンシング反応キット (Applied Biosystems) を用いてシーケンシングした。

20

【0123】

5 cDNAクローンの相同性検索と推定タンパク質

配列表のcDNAまたは推定アミノ酸配列を用いて、GenBank、SwissProt、BLOCKSなどのデータベースに問合せした。BLASTかBLAST2(前出のAltschulら;前出のAltschul) 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)生物のシーケンシングした。別法として、SmithとSmith (1992, Protein Engineering 5:35-51)で説明されているようなアルゴリズムを用いて、一次配列パターンと二次構造ギャップペナルティで処理することが可能であった。この出願で開示されているすべての配列は、少なくとも49ヌクレオチドの長さを有し、不要な塩基は1.2%に過ぎない(A、C、GまたはTよりもNが記録されている)。

30

【0124】

前出のKarlinで詳説されているように、問合せ配列とデータベース配列の間のBLAST一致を統計的に評価して、ヌクレオチドに $10^{-2.5}$ 、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。BLASTのヌクレオチドあるいはアミノ酸の一致率[問い合わせ配列と参照配列間]に%最大限BLASTスコアを乗じ[問い合わせ配列と参照配列の長さに基づく]、次いで100で除する。実験室で用いるハイブリダイゼーション法と比較して、厳密な一致の電子ストリンジェンシーを約40の下限(不要な塩基のために1-2%のエラーがある)から70の100%一致まで設定した。

40

【0125】

無料で入手可能な配列比較アルゴリズムであるBLASTソフトウェア一式(NCBI, Bethesda MD, <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>)には、核酸分子をアラインメントする「blastn」、核酸分子またはアミノ酸分子を対で直接比較するために用いるBLAST 2など多種の配列分析プログラムが含まれる。BLASTプログラムは、一般的には、ギャップ及びデフォルト設定に設定された他のパラメータと共に用いる。例えば: Matrix: BLOSUM62; Reward for match: 1; Penalty for mismatch: -2; Open Gap 5 及び Extension Gap: 2 penalties; Gap x drop-off: 50; Expect: 10; Word Size: 11; 及び Filter: on同一性は配列全体またはより小さいその部分の長さにより測定する。Brenner ら (1998; Proc Natl Acad

50

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。

【0126】

この出願のcDNAをアSEMBLしたコンセンサス配列あるいはLIFESEQ GOLDデータベースで見つけた鋳型と比較した。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、クオリティスコアを指定した。クオリティスコアが許容できるすべての配列を多様なプリプロセッシングとエディティングの経路にかけ、低質の3'末端、ベクターとリンカー配列、ポリAテイル、Alu繰返し、ミトコンドリア及びリボゾーム配列、細菌汚染配列を除去する。編集した配列は少なくとも長さが50 bpであり、ジヌクレオチド反復、Alu反復などの情報に乏しい配列と反復エレメントを"Ns"あるいはマスクしたものに置換した。

10

【0127】

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

20

【0128】

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

30

【0129】

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

40

【0130】

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

50

たKSRCCをコードするcDNAの断片は、すべての関連する調節配列とコード配列を同じ位置に割り当てる。遺伝地図上の位置は、ヒト染色体の範囲または間隔として表される。cM間隔の地図上の位置は、染色体のpアームの末端に関連して測定する(ヒト中のDNAの1メガベース(Mb)にほぼ等しい)。

【0131】

7 ハイブリダイゼーション技術と分析

基質上でのcDNAの固定

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

10

【0132】

2番目の方法では、インサートに隣接するベクター配列に相補的なプライマーを用いて、cDNAを30サイクルのPCRで細菌ベクターから増幅する。PCR増幅で1~2 ngの初期量の核酸から5 µgより大きい最終量まで増大する。SEPHACRYL-400 ビーズ(APB)を用いて長さが400 bpから約5000 bpまで増幅した核酸を精製する。手であるいはドット/スロット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

【0133】

膜ハイブリダイゼーションのプロープ調製

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

40

【0134】

ポリマーコートされたスライドハイブリダイゼーションのプロープ調製

サンプルから分離したmRNAに由来するハイブリダイゼーションのプロープを使用して、アレイベースのハイブリダイゼーションで配列表のcDNAをスクリーニングする。9 µl TE

50

バッファーで濃度200 ngにmRNAを希釈し、5 μ l 5x バッファー、1 μ l 0.1 M DTT、3 μ l Cy3またはCy5標識化混合液、1 μ l RNアーゼ阻害因子、1 μ l 逆転写酵素、5 μ l 1x 酵母調節mRNAを加え、GEMbrightキット (Incyte Genomics)を用いてプローブを調製する。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, 未発表)。定量用対照として、サンプルmRNAに0.002 ng、0.02 ng、0.2 ng、2 ngでの対照mRNAの1つの設定をそれぞれ1:100,000、1:10,000、1:1000、1:100 (w/w)の比で逆転写反応混合液に希釈する。mRNA示差発現パターンを調べるために、対照mRNAの2番目の設定を1:3, 3:1, 1:10, 10:1, 1:25, また25:1 (w/w)の比で逆転写反応混合液に希釈する。反応混合液を混合して、37 $^{\circ}$ C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 $^{\circ}$ C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 5 M 酢酸ナトリウム、300 μ l 100% エタノールを加えて、エタノール析出させた。プローブを20分間、20,800xgで遠心分離する。そしてペレットを12 μ lの再懸濁バッファーで再懸濁し、5分間65 $^{\circ}$ Cに加熱してから、十分に混合する。プローブを加熱して、以前のように混合して、氷上に保管する。以下に詳述するように、プローブを高密度アレイベースのハイブリダイゼーションにおいて使用する。

10

【0135】

膜系ハイブリダイゼーション

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

20

【0136】

ポリマーコートされたスライドハイブリダイゼーション

プローブを5分間65 $^{\circ}$ Cに加熱してから、5415Cマイクロ遠心分離で5分間、9400 rpmで遠心分離する(Eppendorf Scientific, Westbury NY)。そして18 μ lのアリクオットをアレイ表面に置き、カバーガラスで覆う。アレイは、顕微鏡用スライドより僅かに大きい空洞を有する防水チェンバーに移す。チェンバーのコーナーに140 μ lの5 x SSCを加えることにより、チェンバー内部を湿度100%に保持する。このアレイを含むチャンバーを、60 $^{\circ}$ Cで約6.5時間インキュベートする。アッセイを1xSSC, 0.1% SDS中で、45 $^{\circ}$ Cで10分間洗浄し、次に0.1xSSCで、45 $^{\circ}$ Cで10分間の洗浄を3回繰り返した後に、乾燥させた。

30

【0137】

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

40

【0138】

ハイブリダイゼーション複合体は、Cy3の励起のためには488 nm、Cy5の励起のためには632 nmでスペクトル線を発生し得るInnova 70混合ガス10 Wレーザー (Coherent, Santa

50

Clara CA) を備えた顕微鏡で検出する。20×顕微鏡対物レンズ(Nikon, Melville NY) を用いて、アレイ上に励起レーザー光の焦点を当てる。アレイを含むスライドを顕微鏡のコンピュータ制御のX-Yステージに置き、20 μm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、レーザーにより2つの蛍光色素を同時に励起する。発光された光は、波長に基づき分離され、2つの蛍光色素に対応する2つの光電子増倍管検出器(PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) に送られる。アレイと光電子増倍管間に設置された好適なフィルターを用いて、シグナルをフィルターする。用いる蛍光色素の最大発光の波長は、Cy3では565 nm、Cy5では650 nmである。プローブ混合液に加えた酵母対照mRNAによって生み出されるシグナル強度を用いてスキャンの感度を較正する。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。

10

【0139】

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

20

【0140】

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、ラット、猿および犬の配列の積スコアを下記のようにして計算した。BLASTスコアにヌクレオチド一致率を乗じ、積を(2つの配列の短い方の長さの5倍)で除し、短い配列の長さへの100%アラインメントが積スコアを100にする。積スコアは2つの配列間の類似性の程度と配列一致の長さの両方を考慮する。例えば、積スコア40では一致はエラーが1%から2%以内の厳密さになり、少なくとも積スコア70では一致は厳密となる。類似または関連する分子は、通常、積スコアが8から40の間を示す分子を選択して同定する。

30

【0141】

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

40

【0142】

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。これらの分子をOLIGOプライマ解析ソフトウェア(Molecular Biology Insights)を使用して選択した。検出については、実施例7に記載されている。プロモーターが結合するのを妨げることにより転写を阻害するため、相補的分子を最も固有な5'配列に結合し、オープンリーディングフレームの開始コドンの5' UTR上流のヌクレオチドを含むよう設計する。相補的分子は、ゲノム配列(エンハンサーまたはイン

50

トロン等)を含み、ポリメラーゼ、転写因子または調節分子の結合のために十分に開くような二重らせんの能力に影響を及ぼす「三重らせん」塩基対の形成に用いられる。翻訳を阻害するためには、相補的分子を設計して、リボソームが哺乳動物タンパク質をコードする mRNA に結合するのを阻害する。

【0143】

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

10

【0144】

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

【0145】

10 配列、マイクロアレイ試薬、及び使用の選択

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

20

【0146】

マイクロアレイを構築するためには、cDNAは、cDNA挿入断片をフランキングするベクター配列に対するプライマーの相補性を用いることで細菌細胞から増幅された。30サイクルのPCRは、初期量のcDNA、1~2ngから最終量のcDNA5µgより多くなるまで増量された。増幅されたcDNAは、次にSEPHACRYL-400カラム(APB)を用いて精製された。精製されたcDNAは、ポリマーコートされたガラススライド上で固定化された。ガラス顕微鏡のスライド(Corning, Corning N.Y.)は、0.1% SDS及びアセトン中で超音波洗浄され、洗浄前後に多量の蒸留水で洗浄された。スライドガラスは、4%フッ化水素酸(VWR Scientific Products, West Chester PA)中でエッチングし、蒸留水中で広範囲にわたって洗浄し、95%エタノール中の0.05%アミノプロピルシラン(Sigma Aldrich)でコーティングする。コーティングしたスライドガラスは、110°Cのオーブンで硬化させる。米国特許第5,807,522号に記載されている方法を用いて、コーティングしたガラス基板にcDNAを付加する。平均濃度が100ng/µlのcDNA1µlを高速機械装置により開放型キャピラリープリンティングエレメント(open capillary printing element)に充填する。次にこの装置が、スライド毎に約5nlのcDNAを分注する。

30

40

【0147】

マイクロアレイには、STRATALINKER UVクロスリンカー(Stratagene)を用いてUV架橋する。マイクロアレイは、室温において0.2% SDSで1回洗浄し、蒸留水で3回洗浄する。非特異的な結合部位は、リン酸緩衝生理食塩水(Tropix, Inc., Bedford MA)における0.2%カゼイン中で60°Cで30分間マイクロアレイをインキュベートし、その後上述したように0.2% SDS及び蒸留水で洗浄することによってブロックする。

50

【 0 1 4 8 】

1 1 サンプルの調査

組織サンプル

マッチした通常の結腸及び癌性結腸若しくは結腸ポリープの組織サンプルは、Huntsman Cancer Institute (Salt Lake City, Utah)によって提供された。ドナー3311は直腸癌と診断された85歳の男性である。ドナー3753は健康なコロノ組織を有する性別不詳の個体である。ドナー3757は、性別不詳の個体であり、直腸癌と診断された。ドナー3756は、性別不詳の個体であり、直腸癌と診断された。ドナー3755は、性別不詳の個体であり、直腸ポリープと診断された。ドナー3754は、性別不詳の個体であり、直腸ポリープと診断された。ドナー3583は、59歳の男性患者で、右結腸切除、乳頭状結腸ポリープ、過形成結腸ポリープ、及び結腸アデノーマであると診断された。ドナー3649は、性別不詳の86歳の個体であり、観血的で十分に分化したアデノカルチノーマであると診断された。ドナー3647は性別不詳の83歳の個体であって、リンパ節に対する転位を伴う侵襲性で、適度に分化したアデノカルチノーマであると診断された。比較は、同一のドナーから採取した、適合した通常組織と、腫瘍即ちポリープ組織とで行われた。ドナー3983は、性別不詳の23歳の個体であり、腺腫性大腸ポリポーシス症候群によるポリープと診断され、またリンパ節に転位した中程度に分化したアデノカルチノーマであると診断された。

10

【 0 1 4 9 】

1 2 KSRCCの発現

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

20

【 0 1 5 0 】

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

30

【 0 1 5 1 】

1 3 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CCMを精製して、マウスやウサギを免疫化するために使用する。下記のプロトコルを用いて抗体を産生する。或いは、レーザGENEソフトウェア (DNASTAR) を用いてCCMのアミノ酸配列を解析し、免疫原性の高い領域を決定する。通常C-末端付近或いは隣接する親水性領域内で見られる抗原性エピトープを選択して、合成し、抗体を生成するために用いられる。通常は、長さ約15残基のエピトープを、Fmocケミストリを用いるABI 431A ペプチドシンセサイザ (Applied Biosystems) を用いて生成し、N-マレイミドベンゾイル-N-ヒドロキシスクシンイミドエステル (MBS) を用いた反応によってKLH (Sigma-Aldrich) に結合させて、免疫原性を高める。

40

【 0 1 5 2 】

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

【 0 1 5 3 】

50

1.4 特異的抗体を用いる天然タンパク質の精製

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

【0154】

10

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

【0155】

20

1.6 2つのハイブリッドスクリーン

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

30

【0156】

発現したタンパク質とcDNA融合タンパク質間の陽性の相互作用が、個々の陽性コロニーを単離して、30 で1から2日間、SD/-Trp/-Ura液体で成育することにより確認できる。培地のサンプルをSD/-Trp/-Ura培地で培養し、コロニーが出現するまで30 でインキュベートする。サンプルをSD/-Trp/-UraとSD/-His/-Trp/-Uraプレートで複製培養する。ヒスチジンを含まない培地では成育せず、ヒスチジンを含むSDで成育するコロニーは、pLexAプラスミドを失っている。ヒスチジンを必要とするコロニーをSD/Gal/Raf/X-Gal/-Trp/-Uraで成育する。そして白色コロニーを単離して、増殖する。タンパク質と物理的に相互作用するタンパク質をコードするcDNAを含むpB42AD-cDNAプラスミドを酵母細胞から単離して、特徴付ける。

40

【0157】

1.7 CCMアッセイ

50

腸に於けるCCMの局在化は、BoIIらにより示されているように蛍光顕微鏡によって検出される(1993;J Biol Chem 268:12901-12911)。腸組織の切片は、2.5%のパラホルムアルデヒド及び0.1%のグルタルアルデヒドで固定され、CCMに対する抗体と共にインキュベートされる。IPの細胞内分布は、ピオチン化ヤギの抗モルモットIgG(Kirkegaard and Perry Laboratories, Gaithersburg Md.)と共にインキュベートされ、次にストレプトアビジンが蛍光染料テキサスレッド(APB)と混合された。

【0158】

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。

10

【0159】

(表の簡単な説明)

表1は、大腸癌、及びマイクロアレイ解析で決定された通常結腸組織に対する結腸ポリープ中のCCMの異なった発現を示す。第1列は、大腸癌及び結腸ポリープを有する患者より採取した組織サンプルのためにlog₂ DE(病変組織/微視的通常組織)として示された平均的特異発現(DE)値をリストする。第2列は、特異的発現値における共分散のパーセンテージ(CV%)をリストしたものである。第3列は、蛍光性緑ダイCy3でラベルされた顕微鏡的通常サンプルのための組織及びドナー(Dn)をリストしたものである。第4列は、蛍光性赤ダイCy5でラベルされた病変サンプルのための組織及びドナー(Dn)をリストしたものである。

20

【図面の簡単な説明】

【0160】

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

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

30

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

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

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

40

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

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

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

50

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

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

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

【図1 - L】図は、cDNA(SEQ ID NO:2)によってコードされた哺乳動物CCM(SEQ ID NO: 1)を示している。翻訳は、MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA.)によってなされた。 10

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

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

【図1 - O】図は、cDNA(SEQ ID NO:2)によってコードされた哺乳動物CCM(SEQ ID NO: 1)を示している。翻訳は、MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA.)によってなされた。 20

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

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

【図1 - R】図は、cDNA(SEQ ID NO:2)によってコードされた哺乳動物CCM(SEQ ID NO: 1)を示している。翻訳は、MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA.)によってなされた。 30

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

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

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

【図1 - V】図は、cDNA(SEQ ID NO:2)によってコードされた哺乳動物CCM(SEQ ID NO: 1)を示している。翻訳は、MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA.)によってなされた。 40

【0161】

【表1】

表1

mean log2 DE (C5/C3)	CV %	Cy3	Cy5
-1.92	4.03	Human, Colon Pool, Nimi, Dn3753	Human, Colon Tumor, Cancer, Dn3311
-1.87	5.44	Human, Colon Pool, Nimi, Dn3753	Human, Colon Tumor, Cancer, Dn3757
-1.66	0	Human, Colon Pool, Nimi, Dn3753	Human, Colon Tumor, Cancer, Dn3756
-1.61	18.66	Human, Colon Pool, Nimi, Dn3753	Human, Colon, Polyp, Dn3755
-1.41	5.18	Human, Colon Pool, Nimi, Dn3753	Human, Colon, Polyp, Dn3754
-2.29	6.8	Human, Colon Pool, Nimi, mw/AdenoCA, Dn36	Human, Colon Tumor, AdenoCA, Dn3649
-1.04	11.76	Human, Colon Pool, Nimi, mw/Cancer, Dn3547	Human, Colon Tumor, Cancer, Dn3647
-0.84	1.41	Human, Colon Pool, Nimi, mw/Adenoma, Dn35	Human, Colon Tumor, Adenoma, Dn3583
C:18	5.23	Human, Colon, Mucosa, mw/AdenoCA, Dn398	Human, Colon, Polyp, mw/AdenoCA, Dn3983

【 1 - E 】

1251 1260 1269 1278 1287 1296
 CTC AAG AGA GAG CAG CAG ATT GAT GAG AAG GAG CAC ACA CCT GAC ATC GTG AAG
 L K R E Q O I D E K E H T P D I V K
 1305 1314 1323 1332 1341 1350
 CTC TAC GAG AAA TTA CGA CTT TCC ATG GAG AAA GTT GAC CAG AAA CCT CCA GAA
 L Y E K L R L C M E K V D Q K A P E
 1359 1368 1377 1386 1395 1404
 TAC ATC AAG ATG GCA GCA TCA TTA ATG GCT GGG GAG ACA ACC TAC AGT CTG GAA
 Y I R M A A S L N A G E T T Y S L E
 1413 1422 1431 1440 1449 1458
 CAT GCC AGT GAC CTT GGA GTC GAG GAG AAA GTG TAT GAG CTG ATA GAC CCT
 H A S D L R V E V Q K V Y E L I D A
 1467 1476 1485 1494 1503 1512
 TTA AGT AAG AAG ATC TTA ACC TTG GGC TTG AAC CAG GAC CCT CCA CCA CAT CCA
 L S K K I L T L G L N Q D P P P H P

【 1 - F 】

1521 1530 1539 1548 1557 1566
 AGC AAT TTG CGG CTG CAG AGA ATG ATC ACA TAC TCA GCT ACA CTT TTT GTG CAG
 S N L K L Q R H I R Y S A T L F V Q
 1575 1584 1593 1602 1611 1620
 GAA AAG TTG CTT GGT TTG ATG TCA CTG CCA ACC AAA GAA CAG TTT GAG GAA CTG
 E K L L G L M S L P T K E Q F E E L
 1629 1638 1647 1656 1665 1674
 AAA AAG AAA AAG AAG GAG GAA ATG GAG AAG AAG GCC GTG AAG ACA GAA CCT
 K K K R K E E M E R R A V E R Q A
 1683 1692 1701 1710 1719 1728
 GCC CTG GAG TCC CAG CGA AGG CTT GAG GAA AGG CAG AGT GGC CTG GCT TCT CGA
 A L E S Q R R L E E R Q S G L A S R
 1737 1746 1755 1764 1773 1782
 GCG GCC AAC GGG GAG GTG GCA TCT CTC GCG AAG GGC CCT GCC CCG TTG AGA AAG
 A A N G E V A S L R R G P A P L R K

【 1 - G 】

1791 1800 1809 1818 1827 1836
 GCT GAG GGC TGG CTC CCA CTG TCA GGA GGT CAG GGG CAG AAT GAG GAC TCA GAC
 A E G W L P L S G G Q G Q S E D S D
 1845 1854 1863 1872 1881 1890
 CCG CTC CTC CAG CAG ATC CAC AAC ATC ACA TCA TTC ATC AAG CAG GCC AAG GCC
 P L L Q Q I H N I T S F I R Q A K A
 1899 1908 1917 1926 1935 1944
 GCG GGC CGC ATG GAT GAA GNG CGC ACT CTG CAG GAG AAG CTG GCG CAG CTC CAG
 A G R M D E V R T L Q E N L R Q L Q
 1953 1962 1971 1980 1989 1998
 GAC GAG TAT GAC CAG CAG ACA GAG AAG GCC ATC GAG CTC TCC CGG AAG CAG
 D E Y D Q Q Q T E K A I E L S R R Q
 2007 2016 2025 2034 2043 2052
 GCT GAG GAG GAG GAC CTG CAG CCG GAA CAG CTG CAG ATG TTG CPT GAA CGG GAG
 A E E D L Q R E Q L Q M L R E R E

【 1 - H 】

2061 2070 2079 2088 2097 2106
 TTG GAA CGA AAG GAG CAG TTT CGG GTS GCA TCC CTG CAC ACA CCG ACT CGG
 L E R E R E Q F R V A S L H T R T R
 2115 2124 2133 2142 2151 2160
 TCC CTG GAC TTC AGA GAA ATC GGC CCT TTT CAG CTG GAG CCC ACC AGA GAG CCT
 S L D F R E I G P F Q L E P S R E P
 2169 2178 2187 2196 2205 2214
 CCC ACC CAC CTT GCT TAT GCT TTG GAT CTA GGC TCT TCC CCA GTT CCA AGC AGC
 R T H L A Y A L D L G S S P V P S
 2223 2232 2241 2250 2259 2268
 ACA GCT CCC AAG ACC CCT TCA CTT AGC TCA ACT CAA CCC ACC AGA GTG TGG TCT
 T A P K T P S L S S T Q P T R V W S
 2277 2286 2295 2304 2313 2322
 GGG CCC CCA GCC GTT GGC CAG GAG CSC TTA CCC CAG AGC ATG CCA CAG CAA
 G P P A V G Q E R L P Q S S M P Q

【 1 - I 】

2331 2340 2349 2358 2367 2376
 CAT GAG GGG CCC TCC TTA AAC CCC TTT GAT GAG GAA GAC CTC TCC AGC CCC ATG
 H E G P S L N P F D E E D L S S P M
 2385 2394 2403 2412 2421 2430
 GAA GAG GCC ACT ACT GGT CCT GCT GCA GGG GTT TCC TTA GAC CTT TCA GCC
 E E A T T G P A A G V S L D P S A
 2439 2448 2457 2466 2475 2484
 CGC ATC CTG AAA GAG TAC AAT CCT TTC GAG GAA GAG GAC GAG GAG GAA GCA
 R I L K E Y N P F E E D E E E E A
 2493 2502 2511 2520 2529 2538
 GTG GCA GGG AAT CCA TTC ATT CAG CCA GAC AGC CCA GCT COT AAC CCC TTC AGT
 V A G N P F I Q P D S P A F N P F S
 2547 2556 2565 2574 2583 2592
 GAG GAA GAC GAA CMT CCC CAG GAG AGG CTC TCA AGC CCT CTG GTT CTT GAT AGC
 E E D E H P Q Q R L S S P L V P G N

【 1 - J 】

2601 2610 2619 2628 2637 2646
 CCC TTT GAG GAA CCC ACC TGT ATC AAC CCC TTT GAG ATG GAC AGT GAC AGT GGG
 P F E E P T C I N P F E M D S D S G
 2655 2664 2673 2682 2691 2700
 CCA GAG GCT GAG GAG CCC ATA GAG GAA GAG CTC CTC CTG CAG CAG ATC GAT AAC
 P E A E P I E E L L L Q Q I D N
 2709 2718 2727 2736 2745 2754
 ATC AAG GCA TAC ATC TTT GAT GGC AAG GAG TCC GGC CTC GAT GAG GAT GAG
 I K A Y I F D A K Q C G R L D E V E
 2763 2772 2781 2790 2799 2808
 GTG CTG ACA GAG AAT CTG CGG GAG CTG AAG CAC ACC CTG GCC CAG AAG GGG
 V L T E N L R E L K H T L A K Q K G
 2817 2826 2835 2844 2853 2862
 GGC ACT GAC TGA CCA GCA GTG GAG AAG GCA CCT TTG GGC CCA GGG GTC TGG CAG
 G T D

【 1 - K 】

2871 2880 2889 2898 2907 2916
 GAG CCA GTG GAG CAG GAC AGA GGG CAG GCA GGA TGG ATG GGG AAG GTG GCA GGG
 2925 2934 2943 2952 2961 2970
 TGA GAA CTC AGA TCC ACA CAG GTG AGS GGC AGG AAT CTG CTG TTT TGT GTT GCG
 2979 2988 2997 3006 3015 3024
 CAC TTT GAG GTA TTT CCA CTA CAG TTG AAT AAT AAA ATA GAA ACT AGA ACA GGG
 3033 3042 3051 3060 3069 3078
 AGA ATC AGC AAT CAG TTG CAG CTT TTC CTG TTT AAT ATT ACT ATC TTT TGT AAT
 3087 3096 3105 3114 3123 3132
 CGG AGG TTT ACC CCT TTT GAA GGG ACT TTA CMT TTT TAC TAC CGA GAT ATA ACT
 3141 3150 3159 3168 3177 3186
 AAA TCC AGC TCT GTT GGG CCC ACG CCA GAA ATG GCT GCT GTG TAC CTC TTG GGT
 3195 3204 3213 3222 3231 3240
 CCA TTT GCT ACT GCC TNG TCT TGG TTC CTT ATG CAG TAT TAT AGS GCA GCC TTT
 3249 3258 3267 3276 3285 3294
 TTA GAG CCC TTC CTT TAA CCA AGA CAG ABA ABA TAA ATT CCA CTG AGC TCT AAT

【 1 - L 】

3303 3312 3321 3330 3339 3348
 CTG CTC TGA CAG AAG TCC ATC CCT AGT AGS CTG TGA GTT CCA TTT CAC CTG GGG
 3357 3366 3375 3384 3393 3402
 CCG CCT CTC CCC TGC TCT GCA CTT CCT GTC TGT TCA ATG GAA GAG GGA GGT GCT
 3411 3420 3429 3438 3447 3456
 GCT ATG AAG GGG AGA GTT TAA ACC CAG GAG AGC CCA GCA CCT CTC TTT AAG GTG
 3465 3474 3483 3492 3501 3510
 GGG TGA TGG GAA TAT TTC ACC AGG GTC TAT TTT CTC AGT TTA AAT TCT TTT TTG
 3519 3528 3537 3546 3555 3564
 TCT CTT TCA GGA AGT TAA GCT CCC AGT GCA GGG TAT CAA TGT GAA TCT GGT CCT
 3573 3582 3591 3600 3609 3618
 GAG CTT TTT AGA AAA TAA GAG TGG TGG CCG GGC CTC GCT CAC GCC TGT AAT
 3627 3636 3645 3654 3663 3672
 CCG AAC ACT TTG GGA GGC CGA GGC AGG CAG ATC ACA AGG TCA GGA GAT CCA GAC
 3681 3690 3708 3717 3726 3735
 CAC CCT GGC TAA CAC AGT GAA ATC CCG TCT CTA CTA AAA AAT CCA AAA ATT AGC

【 1 - M 】

3735 3744 3753 3762 3771 3780
 CAG GCG TGG TGG TGG GCG CCT GTC CCG GCT ACT TTG GAG GCT GAG GCA GGA
 3789 3798 3807 3816 3825 3834
 GAA TGC TGT GAA CCC GGG AGG CAG ACC TTG CAG TGG GCC GAG ATC GCA CCA CTG
 3843 3852 3861 3870 3879 3888
 CAG TCC AGC CTG GGC AGT GAG ACT CCG TCT GAA AAA AAA ATA AGA GAG
 3897 3906 3915 3924 3933 3942
 GTT AAC TTT TTG GAT ATT AGT ATT ATT TGT GAG AAG AGT TTC CCC ACC CTT
 3951 3960 3969 3978 3987 3996
 CTT CTG ATA GAA CAA TTG TCT GTC ACT GGA GAA ATC TCC CTC CAG AGC TTT GGC
 4005 4014 4023 4032 4041 4050
 AAA GTT ACT CTG ATC TGG GTC TGT TTA AAA GGC CGG GTC TCT AAT TTA GAA ATC
 4059 4068 4077 4086 4095 4104
 GGT GAT TTG TGA GCT TTT GCA GAA CMT CAC CAG AAG AAG GGA AGC TTC CCA GAG
 4113 4122 4131 4140 4149 4158
 TCA GAA GCT AAA TTA AAA TAA TTT CCA AGG CTA TTT GAT CAG CCT TCT TCC TTT

【 1 - N 】

4167 4176 4185 4194 4203 4212
 TTG GTT CAT TGT GTC CTG ACT TGG GGC ACT GAT GAG AAT TTT TAT TTT TGA
 4221 4230 4239 4248 4257 4266
 GAC GGA GTT TCA CTT TGT TGC CCA GGC TGG AGT GTA GTG GCA TGA TCT CCG CTC
 4275 4284 4293 4302 4311 4320
 ACT GCA ACC TCC ACC TCC CTG GTT CMA GTG ATA TTT GTG CCT CAG CCC CCA GAA
 4329 4338 4347 4356 4365 4374
 TAG CTG GAA TTA CAG GCG TGG GCC ACC ACG CCT GGT TAA TCT CAG TAT TGG
 4383 4392 4401 4410 4419 4428
 TAG AGG TAG AGT TTC ACC AAT TTG GCC AGG CTG GGT TCG AAC TCC TGG CCT CAA
 4437 4446 4455 4464 4473 4482
 GTG ATC TGC CAC CTT GGC CTC CCA AAG TCC TGG GAA TAC AAG CBT GAG CCA CTG
 4491 4500 4509 4518 4527 4536
 TGC CCG CCC TGA TGA GAT AAT TTA TTA CCA AAG TTA GTG AGA AAC TGA AAT
 4545 4554 4563 4572 4581 4590
 GTT TGA AGA AGC ACA ACC CAG GAT GGT GCT GGT AGC ACC ACA GTA CTT AAA CTG

【 1 - O 】

4599 4608 4617 4626 4635 4644
 TTG GTC AAT TAA GGC CAG AAA GGG AAA TTG TTA AAT TAG CTC TGG TGC TTT GGT
 4653 4662 4671 4680 4689 4698
 TTA CAG GAA CMT AAC TCT TAA CTG ACA TCT GAC ATC ATG AFA GCC ATA TGT GCT
 4707 4716 4725 4734 4743 4752
 CAG CTC TGG GTA GAG TTT CTG CAG TTA CTC ATC TGA ACT AAT GAA CMA TTA CTG
 4761 4770 4779 4788 4797 4806
 ACC ACT AGT CAC TTT ATG CCG TGT AAC TAG CTC TAG GCC AFA CTT TCA CTG GTT
 4815 4824 4833 4842 4851 4860
 ACT GGT GTC AAA GCT GAA AAT TTT CAT TTT GTC ACA TTC TGG TGA AGA CCC CTC TTG
 4869 4878 4887 4896 4905 4914
 ATA TGA GGA ATG TTT TTA CTC TCT TGA TGA AAA AAT AAT CTG TAT TTG TGT TGA
 4923 4932 4941 4950 4959 4968
 TGT TCA CMT TTC TGT AGC ACA TTT CTT ATC CTT TTG GTT GAA TGA AAA GAT CTT
 4977 4986 4995 5004 5013 5022
 GTA TAG GGG TGT GGA GAT GGG GAG TGG GTA GAA GTG TGT GAA GGA CGC TTT GCA

【 1 - P 】

5031 5040 5049 5058 5067 5076
 TTT GGG ATC TGT TCA CAA ACA GCC AFA TGA GTG TGT TAA TGA AAG TCA GCC AGT
 5085 5094 5103 5112 5121 5130
 TAC CAA CCC TGC TGG TTG TTA TGG GAT GTT TTG AGA AGT TGG CAA CCA GGC ATC
 5139 5148 5157 5166 5175 5184
 TAA GAT GTT GCC TGG TAC AGG CCT CTT TTC CCA GAG GCC CMT GAC AAT TCT
 5193 5202 5211 5220 5229 5238
 CTG TTA CTC TAG AGG GTT TCT CAG ATG GCC AGT AGG CTC CCT GCT TAG GGT
 5247 5256 5265 5274 5283 5292
 CTC AAT TCT CTG AAA AGA GGA TGA ACT GAA AAG CAG GTA GTT CCA GAA GCT AAT
 5301 5310 5319 5328 5337 5346
 TGC TGC TTT CCA TCA TAA TTA TTT TTC TTG TGA GAA CMT TTC TCT TTT AAT TAG
 5355 5364 5373 5382 5391 5400
 CTA GTG AAT TTG AAT AAG ACT AAT TCA CTA AAC AFA CCC TTC CCT CAA ATC ACC
 5409 5418 5427 5436 5445 5454
 TCA GGT AGC AAT CTG TAA GTA ACT AAA AGC AAT GAA AAA CAC ACA AGA AAC AAT

【 1 - Q 】

5463 5472 5481 5490 5499 5508
 TTT AAA AAC TAT TTT TAA AGG CCT GGC CGG GTG CAG TGG CTC ACT CCT GTA ATC

5517 5526 5535 5544 5553 5562
 CCA GCA CTT TGG GAG GCC GAG ACG GGT GGA TCA CCT GAG GTC AGG AAT TTG AGA

5571 5580 5589 5598 5607 5616
 CTA GCA TGG CCA ACA TGG TGA AAC CCC GTG TCT ACT AAA AAT ACA AAA AAT AGC

5625 5634 5643 5652 5661 5670
 TGG GCG TGG TGG CAG AGC CCT GTA ATC CCA GCT ACT CAG GGA GGC TGA GGC ACG

5679 5688 5697 5706 5715 5724
 AGA ATC GCT TGA ACC TGG GAG CCG GAG GTT CCA GTG AGC CAA GAT TGC GCC ACT

5733 5742 5751 5760 5769 5778
 GCA CTC CAA CCT GGG TGA CAA GAG CAA GAC TCC TCA CAA AAA ACA AAC AAA

5787 5796 5805 5814 5823 5832
 ACT GAC TGA TTG AAT ATA CTA AAC CAA ACT AAA ATC ATA TTC TTT TGA TTA AGT

5841 5850 5859 5868 5877 5886
 TTA TCC ATG GCT GTA TTC TTT TTT GAA TTC TTC CMT TTA AAT CTT CCA GBA TTA

【 1 - R 】

5895 5904 5913 5922 5931 5940
 TCT TCT CCT GTA AAT ACA TCA CAG GAG TTA GAA TTC TCT ACC CAT CAG CTG TAC

5949 5958 5967 5976 5985 5994
 CAT GTC CCA GAA TTC ATG CAG GCA CAA AGT TGG AGT TAC AGA GAT GGG TTG ACA

6003 6012 6021 6030 6039 6048
 GCA GGC AAA CTT GGC CTA TGT AAT ATA ACC ACA ACT TCA AGT TCT TAC CTC ATG

6057 6066 6075 6084 6093 6102
 TGA ATA TTC ACC CTT TCT TTA GTC TTC CAA GGC AAA CAG CCC GGT CTC ATC ACC

6111 6120 6129 6138 6147 6156
 AGA TGA GCA AGG TCT TGA TAT GGC ATA GCA GAT CTC CCT AGA CAC AGA TCA TGA

6165 6174 6183 6192 6201 6210
 CAA AAG ATG GAA GAG ACT TAG GGA TTC AGG CAT CAG ATG AAG TTG GCT TTT CCC

6219 6228 6237 6246 6255 6264
 TTT TAT GGC TTG TTT GTA TTT ACC CTG TCT AAT ACA CTA AGG ATA CTT ACT CAT

6273 6282 6291 6300 6309 6318
 TGT ACT TGC AGC TCA ATA TGT CTT TGC TCA GAT ACT AAA ATG TAC CTC TGA

【 1 - S 】

6327 6336 6345 6354 6363 6372
 GTC AAT GTG AGC TGT GTG GTA GGT TGC ACA TTG GCA TAG TTG GTG ATG GGA CTC

6381 6390 6399 6408 6417 6426
 AAA ATG AAA AGG TGG TCT CTT TAC CAG GTC ACA GAC TGT AGC AGA TTG TGC TTG

6435 6444 6453 6462 6471 6480
 TTA TCT GAC AAT GAC TGT CAC TTT GAG GGT CPT TGA TTT GCA TGC ACT ACT CTG

6489 6498 6507 6516 6525 6534
 GGG CPT TPT AAT GGA GCC TTT TTT AAA AAA AAT AAA ATC TGA GAT AGA GGT TGG

6543 6552 6561 6570 6579 6588
 GGG TGT GTG TGT CTG TGT GCA CAT GTG TTA CAA GTG AGA ATC ATC AGA TGA CAT

6597 6606 6615 6624 6633 6642
 CCC TTC TCC TTT CTT CTT GAC AAC CAT CTG AET ATC AGA ATA GTT CCA GCA CCT

6651 6660 6669 6678 6687 6696
 GTG TTG TTT GTC TGG TTA AGG CCT CTG GAA AAA ATG AAG GTC ACT GGG TTC TGA

6705 6714 6723 6732 6741 6750
 ACA GGG GAT AGA TAC GGG TTC CAG TTC TGC CCT TAT TCC CAG TTA TTC CTG CAG

【 1 - T 】

6759 6768 6777 6786 6795 6804
 TGC TGG TTA AAT GAA CAG TTT TAC AGA AGT ATA GNG CAG ATC CCT TAA TGA TCC

6813 6822 6831 6840 6849 6858
 TAT CAT CCC AGT TTC ACA CTG GAA GAA ACT GAG GCT CAG AGA TTG AGG AAC CTT

6867 6876 6885 6894 6903 6912
 GTG ACT GCC ATC TGT GCC AGT CAC CAC CTC TGC ATG ACC ACA TTT CCC TCC ATT

6921 6930 6939 6948 6957 6966
 AGC ACT ACC AGC ATG CCT GTG AGG TNG TTA CTC AGC TGT TCA TCT GAC CCC AGA

6975 6984 6993 7002 7011 7020
 CPT AGA AAA GTT AAG AGA GAT TNG CCT GGC TTA GGC CAC ACA GCC AGT CAG TGG

7029 7038 7047 7056 7065 7074
 TGG AAT GAG GGT TTA AAT CCA GAT CTG CCT GAC TTA AAT TGC CCT GAG TAA TAA

7083 7092 7101 7110 7119 7128
 GTA GCT GGG TCA GGT AGT GGT GTG GTG GGT GCT CAA CCA TTG TGT CAA

7137 7146 7155 7164 7173 7182
 GCT CAA AGA GGG CAG AAA AGG CCA ATG GGG AAT GAG GTT GGG TGC TGC TGA GCT

【 1 - U 】

7191 7200 7209 7218 7227 7236
 GGT TTC CGA TAG TGA CTG CCT GAG TCT GTS AAA CTG AGA CCT CTT TAG GGC CTT

7245 7254 7263 7272 7281 7290
 AAG AAT ATA TAG ATT CTT AAA CCA GSA GAG TAT AAT ATC TAG TGC TCA CTG TAG

7308 7317 7326 7335 7344
 GCT AGA CTG AGT GCT AAG TAT TAC GCA TGG CTT ACC TCA GTC TTC CCA ACA ACC

7353 7362 7371 7380 7389 7398
 CAT TTA CAC AAG CAG AAA TGA GTG AGT GGT GAA GTT AAG CTT GCT TAG GGT TGC

7407 7416 7425 7434 7443 7452
 ACT GCT GGG TTT CCA GCC TAG GTC TGT CTG ACT GCA AAA CCC AAG CTG CCC TAG

7461 7470 7479 7488 7497 7506
 ATT GTG TTG TAT GTG GAC ACC CAT GTT TCT AAT GTG GAA TGG GGT GTT TCA TCC

7515 7524 7533 7542 7551 7560
 CTA GCA AGT GCC TCC TGT GSA CTT TTA AGA TGA AAA ACT TCT GCC TAA CTC TTC

7569 7578 7587 7596 7605 7614
 AAG AAG CCT AGC GTA GTT GTA AAA ATA CTT CAG AAA GCA CTT CTG GTT TTA GCT

【 1 - V 】

7623 7632 7641 7650 7659 7668
 CAG ACA CGT ABA GTT TGA AAG TTA TCC TAT CCT TAC CAC AAG AAA AAG CTG GGC

7677 7686 7695 7704 7713 7722
 AAA CTC AAA AAG ATT GGC CTT TCT TGG TCC CAT TAG AGA GCT GAG TGT GCA GGT

7731 7740 7749 7758 7767 7776
 CAA ACT GCC ATC CCA AAA TCT AGA GAG AGG TGA GTC CAG AGA GTG AGC AAG ATC

7785 7794 7803 7812 7821 7830
 TAA TTG CCT GSA GCA GAA GCC ACG GGA GCC ATA GCT AGT AAG AAA ACT TPA ATG

7839 7848 7857 7866 7875 7884
 GTA AAT TTG ATC AAC TGC TGG AGG TTG AGT GTG GAC TAG CAT GGG AAG GAG CAA

7893 7902 7911 7920 7929 7938
 CTC CCC CTC ACC CCC AAG GCT TTC CTG GGT TTT ACC CTA CTT GGT TCT AAT GGT

7947 7956 7965 7974 7983 7992
 GAA GAG CTG AAG AAG ACC CTC TGG TGA CTC TGG CAG GGG AAA GCG ACA ATC ATC

8001 8010 8019
 TTT GTA ATC AAG GCC TTC TCC ATA ATA AA 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
1 August 2002 (01.08.2002)

PCT

(10) International Publication Number
WO 02/059311 A2

- (51) International Patent Classification: C12N 15/12, BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US).
- (21) International Application Number: PCT/US01/48917 (74) Agents: STREETER, David, G. et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (22) International Filing Date: 13 December 2001 (13.12.2001) (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KH, 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: 09/740,235 18 December 2000 (18.12.2000) US (84) Designated States (regional): ARIPO patent (GH, GM, KI, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CI, CG, CL, CM, GA, GN, GQ, GW, ML, MR, NI, SN, TD, TG).
- (71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (72) Inventors and (75) Inventors/Applicants (for US only): WOLVEN, Amy [US/US]; 237 41st Street, #5, Oakland, CA 94611 (US); KRASNOW, Randi [US/US]; 817 Santa Fe Avenue, Stamford, CA 94305 (US); WARREN, Bridget, A. [US/US]; 10130 Parkwood Drive #2, Cupertino, CA 95014 (US).

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

[Continued on next page]

(54) Title: COLON CANCER MARKER

```

          9       18       27       36       45       54
5'  GCC TGC CGA GGT CCG AGC GAC CGA TGG AGA TGG CGG CTG CGG CTG AGT GAC GGA
          63       72       81       90       99      108
    CGG TGG AGG CCC AGA GCC CGG GCC TGA AGG GGG GGA CAA ACC TGG GTG CCC GCA
          117      126      135      144      153      162
    GGA GCC CGG CAG GGT GTC TTA CAA GTA TCA AGA ACT TAC TAT ATG TGG TTG AAT
          171      180      189      198      207      216
    AAA CAA TCA AGG TAA AGA GCA TCA AGT AAA AAC TTC TGC TTG TTG ATA AGT ACT
          225      234      243      252      261      270
    TCA GAC ATT CCC CCA GTG GCT GAA GTG GCA TAT GAA TTA TGA AGT TGG ATC ATT
          279      288      297      306      315      324
    TGG AAT GAA TGT AAG AGA ATT GCC AAG GGC TCC TCC TAC TCC AGA GAG GAA ACC
          333      342      351      360      369      378
    TCA TCC AGG GCC ATG AAG CCA CTT CCT CAC CAT CTG TGT GCT GCT TAA GCT AAT
          387      396      405      414      423      432
    GCT CGG GGA ACC ATG GTT CCT TGG GAG GAA TCA AGC TGA CTC TTG GCA TGA GAT

```

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

WO 02/059311 A2

WO 02/059311 A2



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

WO 02/059311

PCT/US01/48917

COLON CANCER MARKER**TECHNICAL FIELD**

This invention relates to a mammalian cDNA which encodes a colon cancer marker and to
5 the use of the cDNA and the encoded protein in the diagnosis and treatment of colon disorders,
particularly colon cancer and polyps.

BACKGROUND OF THE INVENTION

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
10 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
15 systems for testing diagnostic and therapeutic agents for human conditions, diseases, and
disorders.

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
20 individuals with specific genetic syndromes (Potter (1999) J Natl Cancer Institute 91:916-932).
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).

Colon cancer is causally related to both genes and the environment. Several molecular
25 pathways have been linked to the development of colon cancer, and the expression of key genes in
any of these pathways may be affected 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.

For example, it is well known that abnormal patterns of DNA methylation occur
30 consistently in human tumors and include, simultaneously, widespread genomic hypomethylation
and localized areas of increased methylation. 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. Indeed, DNA methyltransferase, the enzyme that performs DNA methylation, is
significantly increased in histologically normal mucosa from patients with colon cancer or in the
35 benign polyps that precede cancer. This increase continues during the progression of colonic
neoplasms (El-Deiry *et al.* (1991) Proc Natl Acad Sci USA 88:3470-3474). Increased DNA

WO 02/059311

PCT/US01/48917

methylation occurs in G+C rich areas of genomic DNA termed "CpG islands" that are important for maintenance of an "open" transcriptional conformation around genes, and hypermethylation of these regions results in a "closed" conformation that silences gene transcription. It has been suggested that the silencing or downregulation of differentiation genes by such abnormal

5 methylation of CpG islands may prevent differentiation in immortalized cells (Antequera *et al.* (1990) *Cell* 62:503-514).

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. FAP is characterized by the early development of multiple colorectal adenomas that

10 progress to cancer at a mean age of 44 years. 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. A series of other genetic changes follow activation of the APC- β -catenin-Tcf pathway and accompanies the transition from normal colonic mucosa to metastatic carcinoma. These changes include mutation

15 of the K-Ras proto-oncogene, changes in methylation patterns, and mutation or loss of the tumor suppressor genes p53 and Smad4/DPC4. While the inheritance of a mutated APC gene is a rare event, the loss or mutation of APC and the consequent effects on the APC- β -catenin-Tcf pathway is believed to be central to the majority of colon cancers in the general population.

Hereditary nonpolyposis colorectal cancer (HNPCC) is another inherited autosomal

20 dominant syndrome with a less well defined phenotype than FAP. HNPCC, which accounts for about 2% of colorectal cancer cases, is distinguished by the tendency to early onset of cancer and the development of other cancers, particularly those involving the endometrium, urinary tract, stomach, and biliary system. HNPCC results from the mutation of one or more genes in the DNA mismatch repair (MMR) pathway. Mutations in two human MMR genes, MSH2 and MLH1, are

25 found in a large majority of HNPCC families identified to date. The DNA MMR pathway identifies and repairs errors that result from the activity of DNA polymerase during replication. Furthermore, loss of MMR activity contributes to cancer progression through accumulation of other gene mutations and deletions, such as loss of the BAX gene which controls apoptosis, and the TGF β receptor II gene which controls cell growth. Because of the potential for irreparable

30 damage to DNA in an individual with a DNA MMR defect, progression to carcinoma is more rapid than usual.

Although ulcerative colitis is a minor contributor to colon cancer, affected individuals have about a 20-fold increase in risk for developing cancer. Progression is associated with mutations in the p53 gene which may occur early, appearing even in histologically normal tissue.

35 The progression of the disease from ulcerative colitis to dysplasia/carcinoma without an

WO 02/059311

PCT/US01/48917

intermediate polyp state suggests a high degree of mutagenic activity resulting from the exposure of proliferating cells in the colonic mucosa to the colonic contents.

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
5 hypermethylation of the ER gene (Issa *et al.* (1994) *Nature Genetics* 7:536-540). Introduction and expression of an exogenous ER coding sequence into cultured colon carcinoma cells results in marked suppression of growth. Inhibition of cancer cell invasion depends on the function of the hormone binding domain and the N-terminal zinc finger region of the ER (Platet *et al.* (2000) *Mol Endocrinol* 14:999-1009). Activation of the ER by hormone binding induces transcription of
10 specific target genes and may be linked to the reduction in cancer invasiveness. In the absence of hormone, protein-protein interactions with the zinc finger region may also contribute to the inhibition of cancer cell migration. The connection between loss of the ER protein in colonic epithelial cells and the consequent development of cancer has not been established.

The FYVE-finger proteins play roles in cellular processes such as receptor signaling,
15 vesicular trafficking, and actin-regulated membrane rearrangements (Stenmark and Aasland (1999) *J Cell Science* 112:4175-4183). The FYVE domain is a type of zinc finger that typically contains eight conserved cysteines which bind two Zn²⁺ cations, has conserved glycine and arginine residues, and a basic motif with the consensus sequence R(R/K)HHCR. The FYVE domain binds to phosphoinositides found in specific membranes. The presence of other domains
20 in FYVE-finger proteins may mediate protein-protein interactions with other molecules at the membrane and may be involved in the recruitment of signaling molecules such as small GTPases to membranes at particular cellular locations. In mice with a null mutation in the FYVE finger protein Hrs, embryos showed defects in ventral folding morphogenesis and died *in utero* (Komada and Soriano (1999) *Genes Dev* 13:1475-1485). The embryos developed with their ventral region
25 outside of the yolk sac, had two independent bilateral heart tubes, and no foregut. Mutations in genes that control cellular differentiation and proliferation in the intestine may be linked with colon diseases (Dove *et al.* (1998) *Phil Trans R Soc Lond B* 353:915-923).

Clearly there are a number of genetic alterations associated with colon cancer and with the development and progression of the disease. Particularly, downregulation of expression or
30 deletion of genes potentially provide early indicators of cancer development, and may also be used to monitor disease progression or provide possible therapeutic targets. The specific genes affected in a given case of colon cancer depend on the molecular progression of the disease. Identification of additional genes associated with colon cancer and the precancerous state would provide more reliable diagnostic patterns associated with the development and progression of the disease.

35 The discovery of a mammalian cDNA encoding a colon cancer marker satisfies a need in

WO 02/059311

PCT/US01/48917

the art by providing compositions which are useful in the diagnosis and treatment of colon disorders, particularly colon cancer and polyps.

SUMMARY OF THE INVENTION

5 The invention is based on the discovery of a mammalian cDNA which encodes a colon cancer marker (CCM), which is useful in the diagnosis and treatment of colon disorders, particularly colon cancer and polyps.

The invention provides an isolated mammalian cDNA or a fragment thereof encoding a mammalian 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 active portion of SEQ ID NO:1. The invention also provides an isolated mammalian cDNA or the complement thereof selected from the group consisting of a nucleic acid sequence of SEQ ID NO:2, a fragment of SEQ ID NO:2 selected from SEQ ID NOs:3-22, and a variant having at least 87% identity to the nucleic acid sequence of SEQ ID NO:2 selected from SEQ ID NOs:23-34. The invention additionally provides a composition, a substrate, and a probe comprising the cDNA, or the complement of the cDNA, encoding CCM. The invention further provides a vector containing the cDNA, a host cell containing the vector and a method for using the cDNA to make CCM. In one aspect, the invention provides a substrate containing at least one of the cDNAs. In a second aspect, the invention provides a probe comprising a cDNA which can be used in methods of detection, screening, and purification. In a further aspect, the probe is a single stranded complementary RNA or DNA molecule.

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

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

WO 02/059311

PCT/US01/48917

molecules, peptide nucleic acids, artificial chromosome constructions, peptides, transcription factors, repressors, and regulatory molecules.

The invention provides a purified mammalian protein or a portion thereof selected from the group consisting of an amino acid sequence of SEQ ID NO:1, a variant having 85% identity to
5 the amino acid sequence of SEQ ID NO:1, an antigenic epitope of SEQ ID NO:1, and a biologically active portion of SEQ ID NO:1. The invention also provides a composition comprising the purified protein or a portion thereof in conjunction with a pharmaceutical carrier. The invention further provides a method of using the CCM to treat a subject with colon disorders, particularly colon cancer and polyps comprising administering to a patient in need of such
10 treatment the composition containing the purified protein. The invention still further provides a method for using a protein to screen a library or a plurality of molecules or compounds to identify at least one ligand, the method comprising combining the protein with the molecules or compounds under conditions to allow specific binding and detecting specific binding, thereby identifying a ligand which specifically binds the protein. In one aspect, the molecules or
15 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 colon disorders, particularly colon cancer and polyps.

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

The invention also provides a method of using a mammalian protein to prepare and purify
25 antibodies comprising immunizing a animal with the protein under conditions to elicit an antibody response, isolating animal antibodies, attaching the protein to a substrate, contacting the substrate with isolated 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 colon disorders, particularly colon cancer and polyps. The invention also provides a method of using an antibody to diagnose colon disorders, particularly colon cancer and polyps comprising combining the antibody comparing the quantity of bound antibody to known standards, thereby establishing the presence of colon disorders, particularly colon cancer and
35 polyps. The invention further provides a method of using an antibody to treat colon disorders,

WO 02/059311

PCT/US01/48917

particularly colon cancer and polyps comprising administering to a patient in need of such treatment a pharmaceutical composition comprising the purified antibody.

The invention provides a method for inserting a marker gene into the genomic DNA of a mammal to disrupt the expression of the endogenous polynucleotide. The invention also provides a method for using a cDNA to produce a mammalian model system, the method comprising constructing a vector containing the cDNA selected from SEQ ID NOs:2-34, transforming the vector into an embryonic stem cell, selecting a transformed embryonic stem, microinjecting the transformed embryonic stem cell into a mammalian blastocyst, thereby forming a chimeric blastocyst, transferring the chimeric blastocyst into a pseudopregnant dam, wherein the dam gives birth to a chimeric offspring containing the cDNA in its germ line, and breeding the chimeric mammal to produce a homozygous, mammalian model system.

BRIEF DESCRIPTION OF THE FIGURES AND TABLE

Figures 1A, 1B, 1C, 1D, 1E, 1F, 1G, 1H, 1I, 1J, 1K, 1L, 1M, 1N, 1O, 1P, 1Q, 1R, 1S, 1T, 1U and 1V show the mammalian CCM (SEQ ID NO:1) encoded by the cDNA (SEQ ID NO:2). The translation was produced using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA).

Table 1 shows the differential expression of CCM in colon cancer and colon polyps relative to normal colon tissue as determined by microarray analysis. Column 1 lists the mean differential expression (DE) values presented as log₂ DE (diseased tissue/microscopically normal tissue) for tissue samples from patients with colon cancer and colon polyps. Column 2 lists the percentage covariance (CV%) in differential expression values. Column 3 lists the tissue and patient donor (Dn) for microscopically normal samples labeled with fluorescent green dye Cy3. Column 4 lists the tissue and patient donor (Dn) for diseased samples labeled with fluorescent red dye Cy5.

DESCRIPTION OF THE INVENTION

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

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention

WO 02/059311

PCT/US01/48917

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.

5 **Definitions**

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

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

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

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

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

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

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

35 "Disorder" refers to conditions, diseases or syndromes in which the cDNAs and CCM are

WO 02/059311

PCT/US01/48917

differentially expressed such as colon disorders, particularly colon cancer and polyps.

"Fragment" refers to a chain of consecutive nucleotides from about 200 to about 700 base pairs in length. Fragments may be used in PCR or hybridization technologies to identify related nucleic acid molecules and in binding assays to screen for a ligand. Nucleic acids and their
5 ligands identified in this manner 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'. The degree of complementarity and
10 the use of nucleotide analogs affect the efficiency and stringency of hybridization reactions.

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

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

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

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

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

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

"Purified" refers to any molecule or compound that is separated from its natural
35

WO 02/059311

PCT/US01/48917

environment and is from about 60% free to about 90% free from other components with which it is naturally associated.

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

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

"Similarity" as applied to sequences, refers to the quantification (usually percentage) of nucleotide or residue matches between at least two sequences aligned using a standardized
15 algorithm such as Smith-Waterman alignment (Smith and Waterman (1981) J Mol Biol 147:195-197) or BLAST2 (Altschul *et al.* (1997) Nucleic Acids Res 25:3389-3402). BLAST2 may be used in a standardized and reproducible way to insert gaps in one of the sequences in order to optimize alignment and to achieve a more meaningful comparison between them.

"Substrate" refers to any rigid or semi-rigid support to which cDNAs or proteins are
20 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
25 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,
30 tertiary, or quaternary structure.

THE INVENTION

The invention is based on the discovery of a mammalian cDNA which encodes a colon cancer marker 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 disorders.

35 Nucleic acids encoding CCM of the present invention were first identified in Incyte Clone

WO 02/059311

PCT/US01/48917

5098390 from the mammary epithelial cell cDNA library (EPIMNON05) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:2, was derived from the following overlapping and/or extended nucleic acid sequences (SEQ ID NOs:3-22): Incyte Clones 5098390F6 (EPIMNON05), 8022266J1, 7394066H1, 6309461H1 (NERDITDN03), 7436032H1, 7676140H2, 6923049H1 (PLACFER06), 8216794J1, 6995895H1 (BRAQTDR02), 6889439H1 (BRAITDR03), 70405975D1 (SG0000195), 8117068H1, 747646H1 (BRAITUT01), 7984383H1, 8073132J1, 6744627H1 (BRAFN0T02), 6749637H1 (BRAXNOT03), 5073165F8 (COLCTUT03), 6441214H1 (BRAENOT02), 2436362H1 (BRAVUNT02), and GenBank EST g921283 (SEQ ID NO:35), and genomic sequence GNFL.g8572864_000002_002.edit (SEQ ID NO:36). Table 1 shows the differential expression of CCM in colon cancer and colon polyps relative to normal colon tissue as determined by microarray analysis. **Differential expression (DE) was considered significant if observed to be at least 2.5-fold in at least one patient and at least 2-fold in a majority of patients.** CCM shows reduced expression in tissues from patients with colon cancer relative to microscopically normal tissue from the same donors (Dn3753, Dn3649, and Dn3647). CCM shows reduced expression in tissue from a patient with colon adenoma relative to microscopically normal tissue from the same donor (Dn3583). CCM also shows reduced expression in tissue from a patient with colon polyps relative to microscopically normal tissue from the same donor (Dn3753). Therefore, the cDNA is useful in diagnostic assays for colon disorders, particularly colon cancer and polyps. A fragment thereof the cDNA from about nucleotide 1 to about nucleotide 50 is also useful in diagnostic assays.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1. CCM is 784 amino acids in length and has one potential N-glycosylation site at N465; one potential cyclic AMP- or cyclic GMP-dependent protein kinase phosphorylation site at S230; ten potential casein kinase II phosphorylation sites at S3, S31, S38, T132, T137, S239, T484, S635, S691, and S726; and nine potential protein kinase C phosphorylation sites at S141, S171, S226, S333, S407, S429, T501, S508, and S654. MOTIFS analysis indicates that the region of CCM from C16 through H37 is similar to a C2H2 type zinc finger motif. BLOCKS analysis indicates that the region of CCM from C182 through M191 is similar to a C3HC4 zinc finger (RING finger) motif. Pfam analysis indicates that the region of CCM from V152 through K261 is similar to a FYVE-finger domain (E-value = 3.8×10^{-15}). In particular, CCM has the eight conserved cysteine residues at C163, C166, C179, C182, C187, C190, C252, and C255; the RRHHCR consensus sequence at residues R175 through R179; and the conserved glycine at G183 typical of FYVE-finger proteins. Useful antigenic epitopes extend from N156 to M186, N488 to L525, and A680 to A730, and biologically active portions of CCM extend from C16 through H37 and from V152 through K261. An antibody which specifically

WO 02/059311

PCT/US01/48917

binds CCM is useful in an assay to diagnose colon disorders, particularly colon cancer and polyps.

The invention also encompasses CCM variants. A preferred CCM variant is one which has at least about 80%, or at least about 90%, or even at least about 95% amino acid sequence identity to the CCM amino acid sequence, and which contains at least one functional or structural characteristic of CCM.

Mammalian variants of the cDNA encoding CCM were identified using BLAST2 with default parameters and the ZOOSEQ databases (Incyte Genomics). These preferred variants have from about 87% to about 95% identity to human CCM as shown in the table below. The first column shows the SEQ ID for the human cDNA (SEQ ID_h); the second column, the SEQ ID for the variant cDNAs (SEQ ID_{var}); the third column, the clone number for the variant cDNAs (Clone_{var}); the fourth column, the library name; the fifth column, the alignment of the variant cDNA to the human cDNA (includes the alignment of different regions of the variant cDNA with different regions of the human cDNA in some cases); and the sixth column, the percent identity to the human cDNA.

SEQ ID _h	SEQ ID _{var}	Clone _{var}	Library Name	N _h Alignment	Identity
2	23	702073518T2	RAKITXT09	949-1543	93%
2	24	702549617T1	RACONON05	994-1544	92%
2	25	702448734T1	RASPNON05	1093-1544	93%
2	26	702081830H1	RABPXT01	958-1376	94%
2	27	702053775T1	RAKITXT08	1151-1442	91%
2	28	700883909H1	RAVANOT01	721-975	90%
2	29	700883983H1	RAVANOT01	721-970	90%
2	30	701092129H1	RALUNOT02	1765-1967	88%
2	31	702137890H2	RABPFT07	597-613	87%
2	32	701741076T1	MNBCNON01	6021-6451	95%
2	33	702769948H2	CNLIHOT01	1377-1833	88%
2	34	702245053H1	CNLIHOT01	3494-3542	93%

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

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

The cDNA and fragments thereof (SEQ ID NOs:2-34) may be used in hybridization, amplification, and screening technologies to identify and distinguish among SEQ ID NO:2 and

WO 02/059311

PCT/US01/48917

related molecules in a sample. The mammalian cDNAs may be used to produce transgenic cell lines or organisms which are model systems for human colon disorders, particularly colon cancer and polyps and upon which the toxicity and efficacy of potential therapeutic treatments may be tested. Toxicology studies, clinical trials, and subject/patient treatment profiles may be performed and monitored using the cDNAs, proteins, antibodies and molecules and compounds identified using the cDNAs and proteins of the present invention.

Characterization and Use of the Invention

cDNA Libraries

In a particular embodiment disclosed herein, mRNA was isolated from mammalian cells and tissues using methods which are well known to those skilled in the art and used to prepare the cDNA libraries. The Incyte clones listed above were isolated from mammalian cDNA libraries. Three library preparations representative of the invention are described in the EXAMPLES below. The consensus sequences were chemically and/or electronically assembled from fragments including Incyte clones and extension and/or shotgun sequences using computer programs such as PHRAP (P Green, University of Washington, Seattle WA), and AUTOASSEMBLER application (Applied Biosystems, Foster City CA). Clones, extension and/or shotgun sequences are electronically assembled into clusters and/or master clusters.

Sequencing

Methods for sequencing nucleic acids are well known in the art and may be used to practice any of the embodiments of the invention. These methods employ enzymes such as the Klenow fragment of DNA polymerase I, SEQUENASE, Taq DNA polymerase and thermostable T7 DNA polymerase (Amersham Pharmacia Biotech (APB), Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 system (Hamilton, Reno NV) and the DNA ENGINE thermal cycler (MJ Research, Watertown MA). Machines commonly used for sequencing include the ABI PRISM 3700, 377 or 373 DNA sequencing systems (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (APB), and the like. The sequences may be analyzed using a variety of algorithms well known 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

WO 02/059311

PCT/US01/48917

reassembled using overlapping ends until the entire sequence of the original insert is known. Shotgun sequencing methods are well known in the art and use thermostable DNA polymerases, heat-labile DNA polymerases, and primers chosen from representative regions flanking the cDNAs of interest. Incomplete assembled sequences are inspected for identity using various algorithms or programs such as CONSED (Gordon (1998) Genome Res 8:195-202) which are well known in the art. Contaminating sequences including vector or chimeric sequences or deleted sequences can be removed or restored, respectively, organizing the incomplete assembled sequences into finished sequences.

Extension of a Nucleic Acid Sequence

10 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

20 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 CCM, allelic variants, or related molecules. The probe may be DNA or RNA, may be single stranded and should have at least 50% sequence identity to any of the nucleic acid sequences, SEQ ID NOs:2-34. 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.

30 The stringency of hybridization is determined by G+C content of the probe, salt concentration, and temperature. In particular, stringency can be increased by reducing the concentration of salt or raising the hybridization temperature. In solutions used for some membrane based hybridizations, addition of an organic solvent such as formamide allows the reaction to occur at a lower temperature. Hybridization can be performed at low stringency with buffers, such as 5xSSC with 1% sodium dodecyl sulfate (SDS) at 60C, which permits the

WO 02/059311

PCT/US01/48917

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

Arrays may be prepared and analyzed using methods known in the art. Oligonucleotides may be used as either probes or targets in an array. The array can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and single nucleotide polymorphisms. Such information may be used to determine gene function; to understand the genetic basis of a condition, disease, or disorder; to diagnose a condition, disease, or disorder; and to develop and monitor the activities of therapeutic agents. (See, e.g., Brennan *et al.* (1995) USPN 5,474,796; Schena *et al.* (1996) Proc Natl Acad Sci 93:10614-10619; Baldeschweiler *et al.* (1995) PCT application WO95/251116; Shalon *et al.* (1995) PCT application WO95/35505; Heller *et al.* (1997) Proc Natl Acad Sci 94:2150-2155; and Heller *et al.* (1997) USPN 5,605,662.)

Hybridization probes are also useful in mapping the naturally occurring genomic sequence. The probes may be hybridized to: 1) a particular chromosome, 2) a specific region of a chromosome, or 3) an artificial chromosome construction such as human artificial chromosome (HAC), yeast artificial chromosome (YAC), bacterial artificial chromosome (BAC), bacterial P1 construction, or single chromosome cDNA libraries.

Expression

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

WO 02/059311

PCT/US01/48917

elements are combined using *in vitro* recombinant DNA techniques, synthetic techniques, and/or *in vivo* genetic recombination techniques well known in the art and described in Sambrook (*supra*, ch. 4, 8, 16 and 17).

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

Routine cloning, subcloning, and propagation of nucleic acid sequences can be achieved using the multifunctional PBLUESCRIPT vector (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Introduction of a nucleic acid sequence into the multiple cloning site of these vectors disrupts the lacZ gene and allows colorimetric screening for transformed bacteria. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence.

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

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

WO 02/059311

PCT/US01/48917

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 thereof may be substantially purified by preparative high performance liquid chromatography and its composition confirmed by amino acid analysis or by sequencing (Creighton (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY).

Preparation and Screening of Antibodies

Various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with CCM or any portion thereof. Adjuvants such as Freund's, mineral gels, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil

WO 02/059311

PCT/US01/48917

emulsions, keyhole limpet hemacyanin (KLH), and dinitrophenol may be used to increase immunological response. The oligopeptide, peptide, or portion of protein used to induce antibodies should consist of at least about five amino acids, more preferably ten amino acids, which are identical to a portion of the natural protein. Oligopeptides may be fused with proteins such as KLH in order to produce antibodies to the chimeric molecule.

5 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. (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 the production of single chain antibodies may be adapted, using methods known in the art, to produce epitope specific single chain antibodies. Antibody fragments which contain specific binding sites for epitopes of the protein may also be generated. For example, such fragments include, but are not limited to, F(ab)² fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab)² fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse *et al.* (1989) *Science* 246:1275-1281.)

20 The CCM 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 involve the measurement of complex formation between the protein and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes is preferred, but a competitive binding assay may also be employed (Pound (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

Labeling of Molecules for Assay

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

WO 02/059311

PCT/US01/48917

amines, thiols and other groups present in the molecules using reagents such as BIODIPY or FITC (Molecular Probes, Eugene OR).

DIAGNOSTICS

The cDNAs, fragments, oligonucleotides, complementary RNA and DNA molecules, and
5 PNAs and may be used to detect and quantify differential gene expression, absence/presence vs. excess, expression of mRNAs or to monitor mRNA levels during therapeutic intervention. Similarly antibodies which specifically bind CCM may be used to quantitate the protein. Disorders associated with differential expression include colon disorders, particularly colon cancer and polyps. The diagnostic assay may use hybridization or amplification technology to
10 compare gene expression in a biological sample from a patient to standard samples in order to detect differential gene expression. Qualitative or quantitative methods for this comparison are well known in the art.

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

Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies and in clinical trial or to monitor the treatment of an individual patient.
30 Once the presence of a condition is established and a treatment protocol is initiated, diagnostic assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in a normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

35 Immunological Methods

WO 02/059311

PCT/US01/48917

Detection and quantification of a protein using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes is preferred, but a competitive binding assay may be employed. (See, e.g., Coligan *et al.* (1997) *Current Protocols in Immunology*, Wiley-Interscience, New York NY; and Pound, *supra*).

THERAPEUTICS

Regions of CCM (SEQ ID NO:1) show chemical and structural similarity to a FYVE-finger domain-containing protein. In addition, differential expression of CCM is highly associated with colon disorders, particularly colon cancer and polyps, as shown in Table 1. CCM clearly plays a role in colon disorders, particularly colon cancer and polyps.

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

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

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

WO 02/059311

PCT/US01/48917

Modification of Gene Expression Using Nucleic Acids

Gene expression may be modified by designing complementary or antisense molecules (DNA, RNA, or PNA) to the control, 5', 3', or other regulatory regions of the gene encoding CCM. Oligonucleotides designed with reference to the transcription initiation site are preferred.

- 5 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 or fragments thereof
- 10 may be screened to identify those which specifically bind a regulatory, nontranslated sequence.

- Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA followed by endonucleolytic cleavage at sites such as GUA, GUU, and GUC. Once such sites are identified, an oligonucleotide with the same
- 15 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.

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

Screening and Purification Assays

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

WO 02/059311

PCT/US01/48917

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 an appropriate chaotropic agent to separate the protein from the purified ligand.

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

In one aspect, this invention contemplates a method for high throughput screening using very small assay volumes and very small amounts of test compound as described in USPN 5,876,946, incorporated herein by reference. This method is used to screen large numbers of molecules and compounds via specific binding. In another aspect, this invention also

WO 02/059311

PCT/US01/48917

contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of binding the protein specifically compete with a test compound capable of binding to the protein or oligopeptide or portion thereof. Molecules or compounds identified by screening may be used in a mammalian model system to evaluate their toxicity, diagnostic, or therapeutic potential.

Pharmacology

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

A therapeutically effective dose refers to that amount of protein or inhibitor which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity of such agents may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., 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 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

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

Toxicology

Toxicology is the study of the effects of agents on living systems. The majority of toxicity studies are performed on rats or mice. Observation of qualitative and quantitative changes in physiology, behavior, homeostatic processes, and lethality in the rats or mice are used

WO 02/059311

PCT/US01/48917

to generate a toxicity profile and to assess potential consequences on human health following exposure to the agent.

Genetic toxicology identifies and analyzes the effect of an agent on the rate of endogenous, spontaneous, and induced genetic mutations. Genotoxic agents usually have common chemical or physical properties that facilitate interaction with nucleic acids and are most harmful when chromosomal aberrations are transmitted to progeny. Toxicological studies identify agents that increase the frequency of structural or functional abnormalities in the tissues of the progeny if administered to either parent before conception, to the mother during pregnancy, or to the developing organism. Mice and rats are most frequently used in these tests because their short reproductive cycle allows the production of the numbers of organisms needed to satisfy statistical requirements.

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

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

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

25 Transgenic Animal Models

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

Embryonic Stem Cells

Embryonic (ES) stem cells isolated from rodent embryos retain the potential to form embryonic tissues. When ES cells are placed inside a carrier embryo, they resume normal

WO 02/059311

PCT/US01/48917

development and contribute to tissues of the live-born animal. ES cells are the preferred cells used in the creation of experimental knockout and knockin rodent strains. Mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and are grown under culture conditions well known in the art. Vectors used to produce a transgenic strain contain a disease gene candidate and a marker gene, the latter serves to identify the presence of the introduced disease gene. The vector is transformed into ES cells by methods well known in the art, and transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains.

ES cells derived from human blastocysts may be manipulated *in vitro* to differentiate into at least eight separate cell lineages. These lineages are used to study the differentiation of various cell types and tissues *in vitro*, and they include endoderm, mesoderm, and ectodermal cell types which differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes.

15 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 blastulae are implanted into pseudopregnant dams. Transgenic progeny are crossbred to obtain homozygous inbred lines which lack a functional copy of the mammalian gene. In one example, the mammalian gene is a human gene.

Knockin Analysis

25 ES cells can be used to create knockin humanized animals (pigs) or transgenic animal models (mice or rats) of human diseases. With knockin technology, a region of a human gene is injected into animal ES cells, and the human sequence integrates into the animal cell genome. Transformed cells are injected into blastulae and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of the analogous human condition. These methods have been used to model several human diseases.

Non-Human Primate Model

35 The field of animal testing deals with data and methodology from basic sciences such as physiology, genetics, chemistry, pharmacology and statistics. These data are paramount in evaluating the effects of therapeutic agents on non-human primates as they can be related to

WO 02/059311

PCT/US01/48917

human health. Monkeys are used as human surrogates in vaccine and drug evaluations, and their responses are relevant to human exposures under similar conditions. Cynomolgus and Rhesus monkeys (*Macaca fascicularis* and *Macaca mulatta*, respectively) and Common Marmosets (*Callithrix jacchus*) are the most common non-human primates (NHPs) used in these investigations. Since great cost is associated with developing and maintaining a colony of NHPs, early research and toxicological studies are usually carried out in rodent models. In studies using behavioral measures such as drug addiction, NHPs are the first choice test animal. In addition, NHPs and individual humans exhibit differential sensitivities to many drugs and toxins and can be classified as a range of phenotypes from "extensive metabolizers" to "poor metabolizers" of these agents.

In additional embodiments, the cDNAs which encode the 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.

15 **EXAMPLES**

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention. For purposes of example, preparation of the human brain posterior parietal cortex (BRAENOT02) and normalized mammary epithelial cell (EPIMNON05) libraries will be described.

20 **I cDNA Library Construction**

Brain Posterior Parietal Cortex Library

The BRAENOT02 cDNA library was constructed from brain posterior parietal cortex tissue removed from the brain of a 35-year-old Caucasian male. The frozen tissue was homogenized and lysed in TRIZOL reagent (0.8g tissue/12 ml; Life Technologies) using a POLYTRON homogenizer (Brinkmann Instruments, Westbury NJ). After brief incubation on ice, chloroform was added (1:5 v/v), and the mixture was centrifuged to separate the phases. The upper aqueous phase was removed to a fresh tube, and isopropanol was added to precipitate RNA. The RNA was resuspended in RNase-free water and treated with DNase. The RNA was re-extracted with acid phenol-chloroform and reprecipitated with sodium acetate and ethanol. Poly(A⁺) RNA was isolated using the OLIGOTEX kit (Qiagen, Chatsworth CA).

The mRNA was handled according to the recommended protocols in the SUPERScript 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

WO 02/059311

PCT/US01/48917

400 bp were ligated into pINCY plasmid (Incyte Genomics). The plasmid pINCY was subsequently transformed into DH5 α competent cells (Life Technologies).

Normalized Mammary Epithelial Cell Library

For purposes of example, the normalization of the mammary epithelial cell library (EPIMNON05) is described. About 3.28×10^6 independent clones of the EPIMNOT01 plasmid library in *E. coli* strain DH12S competent cells (Life Technologies) were grown in liquid culture under carbenicillin (25 mg/l) and methicillin (1 mg/ml) selection following transformation by electroporation. To reduce the number of excess cDNA copies according to their abundance levels in the library, the cDNA library was normalized in a single round according to the procedure of Soares *et al.* (1994, Proc Natl Acad Sci 91:9228-9232), with the following modifications. The primer to template ratio in the primer extension reaction was increased from 2:1 to 10:1. The dNTP concentration in the reaction was reduced to 150 μ M for each dNTP to allow the generation of longer (400 to 1000 nt) primer extension products. The reannealing hybridization was extended from 13 to 48 hr. The single stranded DNA circles of the normalized library were purified by hydroxyapatite chromatography and converted to partially double-stranded by random priming, ligated into pINCY plasmid and electroporated into DH12S 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 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 produced by the bacteria (pSPORT 1-ARI) showed no digestion with EcoRI and was digested with Hind III (New England Biolabs) and the overhanging ends were again filled in with Klenow and dNTPs. A linker sequence was phosphorylated, ligated onto the 5' blunt end, digested with EcoRI, and self-ligated. Following transformation into JM109 host cells, plasmids were isolated and tested for preferential digestibility with EcoRI, but not with Hind III. A single colony that met this criteria was designated pINCY plasmid.

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

III Isolation and Sequencing of cDNA Clones

WO 02/059311

PCT/US01/48917

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

10 The cDNAs were prepared for sequencing using the MICROLAB 2200 system (Hamilton) in combination with the DNA ENGINE thermal cyclers (MJ Research). The cDNAs were sequenced by the method of Sanger and Coulson (1975; J Mol Biol 94:441-448) using an ABI PRISM 377 sequencing system (Applied Biosystems) or the MEGABACE 1000 DNA sequencing system (APB). Most of the isolates were sequenced according to standard ABI protocols and kits
15 (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

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'
20 extension of the known fragment. The initial primers were designed using OLIGO primer analysis software (Molecular Biology Insights), to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68C to about 72C. Any stretch of nucleotides that would result in hairpin structures and primer-primer dimerizations was avoided.

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

30 High fidelity amplification was obtained by PCR using methods such as that taught in USPN 5,932,451. PCR was performed in 96-well plates using the DNA ENGINE thermal cycler (MJ Research). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+} , $(NH_4)_2SO_4$, and β -mercaptoethanol, Taq DNA polymerase (APB), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the
35 following parameters for primer pair PCI A and PCI B (Incyte Genomics): Step 1: 94C, three min;

WO 02/059311

PCT/US01/48917

Step 2: 94C, 15 sec; Step 3: 60C, one min; Step 4: 68C, two min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68C, five min; Step 7: storage at 4C. In the alternative, the parameters for primer pair T7 and SK+ (Stratagene) were as follows: Step 1: 94C, three min; Step 2: 94C, 15 sec; Step 3: 57C, one min; Step 4: 68C, two min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68C, five min; Step 7: storage at 4C.

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

The extended clones were desalted, concentrated, transferred to 384-well plates, digested with CviII cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC18 vector (APB). For shotgun sequences, the digested nucleotide sequences were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and the agar was digested with 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.

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

V Homology Searching of cDNA Clones and Their Deduced Proteins

The cDNAs of the Sequence Listing or their deduced amino acid sequences were used to query databases such as GenBank, SwissProt, BLOCKS, and the like. These databases that contain previously identified and annotated sequences or domains were searched using BLAST or

WO 02/059311

PCT/US01/48917

BLAST 2 (Altschul et al. *supra*; Altschul, *supra*) to produce alignments and to determine which sequences were exact matches or homologs. The alignments were to sequences of prokaryotic (bacterial) or eukaryotic (animal, fungal, or plant) origin. Alternatively, algorithms such as the one described in Smith and Smith (1992, Protein Engineering 5:35-51) could have been used to deal with primary sequence patterns and secondary structure gap penalties. All of the sequences disclosed in this application have lengths of at least 49 nucleotides, and no more than 12% uncalled bases (where N is recorded rather than A, C, G, or T).

As detailed in Karlin (*supra*), BLAST matches between a query sequence and a database sequence were evaluated statistically and only reported when they satisfied the threshold of 10^{25} for nucleotides and 10^{14} for peptides. Homology was also evaluated by product score calculated as follows: the % nucleotide or amino acid identity [between the query and reference sequences] in BLAST is multiplied by the % maximum possible BLAST score [based on the lengths of query and reference sequences] and then divided by 100. In comparison with hybridization procedures used in the laboratory, the electronic stringency for an exact match was set at 70, and the conservative lower limit for an exact match was set at approximately 40 (with 1-2% error due to uncalled bases).

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

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

WO 02/059311

PCT/US01/48917

"Ns" or masked.

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

Bins were compared to one another and those having local similarity of at least 82% were combined and reassembled. Bins having templates with less than 95% local identity were split. Templates were subjected to analysis by STITCHER/EXON MAPPER algorithms that analyze the probabilities of the presence of splice variants, alternatively spliced exons, splice junctions, differential expression of alternative spliced genes across tissue types or disease states, and the like. Assembly procedures were repeated periodically, and templates were annotated using BLAST against GenBank databases such as GBpri. An exact match was defined as having from 95% local identity over 200 base pairs through 100% local identity over 100 base pairs and a homolog match as having an E-value (or probability score) of $\leq 1 \times 10^{-8}$. The templates were also subjected to frameshift FASTx against GENPEPT, and homolog match was defined as having an E-value of $\leq 1 \times 10^{-8}$. Template analysis and assembly was described in USSN 09/276,534, filed March 25, 1999.

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

VI Chromosome Mapping

Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR),

WO 02/059311

PCT/US01/48917

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

VII Hybridization Technologies and Analyses

Immobilization of cDNAs on a Substrate

The cDNAs are applied to a substrate by one of the following methods. A mixture of cDNAs is fractionated by gel electrophoresis and transferred to a nylon membrane by capillary transfer. Alternatively, the cDNAs are individually ligated to a vector and inserted into bacterial host cells to form a library. The cDNAs are then arranged on a substrate by one of the following methods. In the first method, bacterial cells containing individual clones are robotically picked and arranged on a nylon membrane. The membrane is placed on LB agar containing selective agent (carbenicillin, kanamycin, ampicillin, or chloramphenicol depending on the vector used) and incubated at 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 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 (APB). Purified nucleic acids are arranged on a nylon membrane manually or using a dot/slot blotting manifold and suction device and are immobilized by denaturation, neutralization, and UV irradiation as described above. Purified nucleic acids are robotically arranged and immobilized on polymer-coated glass slides using the procedure described in USPN 5,807,522. Polymer-coated slides are prepared by cleaning glass microscope slides (Corning, Acton MA) by ultrasound in 0.1% SDS and acetone, etching in 4% hydrofluoric acid (VWR Scientific Products, West Chester PA), coating with 0.05% aminopropyl silane (Sigma Aldrich) in 95% ethanol, and curing in a 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 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

WO 02/059311

PCT/US01/48917

phosphate buffered saline (PBS; Tropix, Bedford MA) for 30 min at 60C; then the arrays are washed in 0.2% SDS and rinsed in distilled water as before.

Probe Preparation for Membrane Hybridization

Hybridization probes derived from the cDNAs of the Sequence Listing are employed for 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

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

Membranes are pre-hybridized in hybridization solution containing 1% Sarkosyl and 1x high phosphate buffer (0.5 M NaCl, 0.1 M Na₂HPO₄, 5 mM EDTA, pH 7) at 55C for two hr. The

WO 02/059311

PCT/US01/48917

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

Probe is heated to 65C for five min, centrifuged five min at 9400 rpm in a 5415C microcentrifuge (Eppendorf Scientific, Westbury NY), and then 18 μ l is aliquoted onto the array surface and covered with a coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μ l of 5xSSC in a corner of the chamber. The chamber containing the arrays is 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 substantially equal numbers of probes derived from both biological samples give a distinct combined fluorescence (Shalon WO95/35505).

Hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective with a resolution of 20 micrometers. In the differential hybridization format, the two fluorophores are sequentially excited by the laser. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. The sensitivity of the scans is calibrated using the signal

WO 02/059311

PCT/US01/48917

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.

5 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
10 overlapping emission spectra) between the fluorophores using the emission spectrum for each fluorophore. A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS program (Incyte Genomics).

15 VIII Electronic Analysis

BLAST was used to search for identical or related molecules in the GenBank or LIFESEQ databases (Incyte Genomics). The product score for human, rat, monkey, and dog 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
20 over the length of the shorter sequence gives a product score of 100. The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and with a product score of at least 70, the match will be exact. Similar or related molecules are usually identified by selecting those which show product scores between 8 and 40.

25 Electronic northern analysis is performed at a product score of 70. All sequences and cDNA libraries in the LIFESEQ database are categorized by system, organ/tissue and cell type. The categories include 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
30 system, sense organs, skin, stomatognathic system, unclassified/mixed, and the urinary tract. For each category, the number of libraries in which the sequence is expressed are counted and shown over the total number of libraries in that category. In a non-normalized library, expression levels of two or more are significant.

IX Complementary Molecules

35 Molecules complementary to the cDNA, from about 5 (PNA) to about 5000 bp

WO 02/059311

PCT/US01/48917

(complement of a cDNA insert), are used to detect or inhibit gene expression. These molecules are selected using OLIGO primer analysis software (Molecular Biology Insights). Detection is described in Example VII. To inhibit transcription by preventing promoter binding, the complementary molecule is designed to bind to the most unique 5' sequence and includes nucleotides of the 5' UTR upstream of the initiation codon of the open reading frame.

Complementary molecules include genomic sequences (such as enhancers or introns) and are used in "triple helix" base pairing to compromise 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.

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

Stable transformation of appropriate dividing cells with a vector encoding the complementary molecule produces a transgenic cell line, tissue, or organism (USPN 4,736,866). 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.

X Selection of Sequences, Microarray Preparation and Use

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

To construct microarrays, cDNAs were amplified from bacterial cells using primers complementary to vector sequences flanking the cDNA insert. Thirty cycles of PCR increased the

WO 02/059311

PCT/US01/48917

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

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

XI Preparation of Samples

Tissue Samples

Matched normal colon and cancerous colon or colon polyp tissue samples were provided by the Huntsman Cancer Institute, (Salt Lake City, UT). Donor 3311 is an 85 year-old male diagnosed with colon cancer. Donor 3753 is an individual, sex unknown, with normal colon tissue. Donor 3757 is an individual, sex unknown, diagnosed with colon cancer. Donor 3756 is an individual, sex unknown, diagnosed with colon cancer. Donor 3755 is an individual, sex unknown, diagnosed with a colon polyp. Donor 3754 is an individual, sex unknown, diagnosed with a colon polyp. Donor 3583 is a 59 year-old male diagnosed with a right colectomy, tubulovillous colon polyp, hyperplastic colon polyp, and colon adenoma. Donor 3649 is an 86 year-old individual, sex unknown, diagnosed with an invasive, well-differentiated adenocarcinoma. Donor 3647 is an 83 year-old individual, sex unknown, diagnosed with an invasive, moderately well-differentiated adenocarcinoma with metastases to the lymph nodes. Comparisons were done with matched normal and tumor or polyp tissue from the same donor. Donor 3983 is a 23 year-old individual, sex unknown, diagnosed with a polyp from adenomatous polyposis coli and with moderately differentiated adenocarcinoma that had metastasized to the lymph nodes.

XII Expression of CCM

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

WO 02/059311

PCT/US01/48917

(Invitrogen, Carlsbad CA) is used to express CCM in CHO cells. The vector contains the selectable *bsd* gene, multiple cloning sites, the promoter/enhancer sequence from the human ubiquitin C gene, a C-terminal V5 epitope for antibody detection with anti-V5 antibodies, and a C-terminal polyhistidine (6xHis) sequence for rapid purification on PROBOND resin (Invitrogen).

- 5 Transformed cells are selected on media containing blasticidin.

Spodoptera frugiperda (Sf9) insect cells are infected with recombinant *Autographica californica* nuclear polyhedrosis virus (baculovirus). The polyhedrin gene is replaced with the cDNA 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

- 10 above. Purified protein is used in the following activity and to make antibodies

XIII Production of Antibodies

CCM 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 CCM is analyzed using LASERGENE software (DNASTAR) to determine regions of

15 high antigenicity. An antigenic epitope, usually found near the C-terminus or in a hydrophilic region is selected, synthesized, and used to raise antibodies. Typically, epitopes of about 15 residues in length are produced using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc-chemistry and coupled to KLH (Sigma-Aldrich) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester to increase antigenicity.

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

XIV 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

30 covalently coupling the antibody to CNBr-activated SEPHAROSE resin (APB). Media containing the protein is passed over the immunoaffinity column, and the column is washed using high ionic strength buffers in the presence of detergent to allow preferential absorbance of the protein. After coupling, the protein is eluted from the column using a buffer of pH 2-3 or a high concentration of urea or thiocyanate ion to disrupt antibody/protein binding, and the protein is collected.

- 35 XV Screening Molecules for Specific Binding with the cDNA or Protein

WO 02/059311

PCT/US01/48917

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

10 XVI Two-Hybrid Screen

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

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

WO 02/059311

PCT/US01/48917

XVII CCM Assay

The localization of CCM in the intestine is detected by fluorescence microscopy as described by Boll *et al.* (1993; J Biol Chem 268:12901-12911). Sections of intestinal tissue are fixed with 2.5% paraformaldehyde and 0.1% glutaraldehyde and incubated with antibodies against CCM. Subcellular distributions of CCM are visualized by incubation with biotinylated goat anti-guinea pig IgG (Kirkegaard and Perry Laboratories, Gaithersburg MD) followed by streptavidin complexed with the fluorescent dye Texas Red (APB).

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 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.

WO 02/059311

PCT/US01/48917

mean log2 DE (C ₅ /C ₃)	CV%	C ₃	C ₅
-1.92	4.03	Human, Colon Pool, Nimi, Dn3753	Human, Colon Tumor, Cancer, Dn3311
-1.87	5.44	Human, Colon Pool, Nimi, Dn3753	Human, Colon Tumor, Cancer, Dn3757
-1.66	0	Human, Colon Pool, Nimi, Dn3753	Human, Colon Tumor, Cancer, Dn3756
-1.61	15.65	Human, Colon Pool, Nimi, Dn3753	Human, Colon Polyp, Dn3755
-1.41	5.18	Human, Colon Pool, Nimi, Dn3753	Human, Colon Polyp, Dn3754
-2.29	6.8	Human, Colon Pool, Nimi, mw/AdenoCA, Dn36	Human, Colon Tumor, AdenoCA, Dn3649
-1.04	11.76	Human, Colon Pool, Nimi, mw/Cancer, Dn3647	Human, Colon Tumor, Cancer, Dn3647
-0.84	1.41	Human, Colon Pool, Nimi, mw/AdenoCA, Dn36	Human, Colon Tumor, AdenoCA, Dn3683
0.16	5.23	Human, Colon, Meeser, mw/AdenoCA, Dn398	Human, Colon Polyp, mw/AdenoCA, Dn3983

TABLE 1

WO 02/059311

PCT/US01/48917

What is claimed is:

1. An isolated cDNA encoding a protein having the amino acid sequence of SEQ ID NO:1.
2. An isolated cDNA selected from:
 - a) a nucleic acid sequence of SEQ ID NO:2 or the complement thereof;
 - 5 b) a fragment of SEQ ID NO:2 selected from SEQ ID NOs:3-22 or the complement thereof;
and
 - c) a variant of SEQ ID NO:2 selected from SEQ ID NOs:23-34.
3. A composition comprising the cDNA or the complement of the cDNA of claim 1.
4. A vector comprising the cDNA of claim 1.
- 10 5. A host cell comprising the vector of claim 4.
6. A method for using a cDNA to produce a protein, the method comprising:
 - a) culturing the host cell of claim 5 under conditions for protein expression; and
 - b) recovering the protein from the host cell culture.
7. A method for using a cDNA to detect expression of a nucleic acid in a sample comprising:
 - 15 a) hybridizing the composition of claim 3 to nucleic acids of the sample, thereby forming
hybridization complexes; and
 - b) comparing hybridization complex formation with a standard, wherein the comparison
indicates expression of the cDNA in the sample.
8. The method of claim 7 further comprising amplifying the nucleic acids of the sample prior to
20 hybridization.
9. The method of claim 7 wherein the composition is attached to a substrate.
10. The method of claim 7 wherein the cDNA is differentially expressed when compared with
the standard and is diagnostic of a colon cancer or colon polyps.
11. A method of using a cDNA to screen a plurality of molecules or compounds, the method
25 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.
- 30 12. The method of claim 11 wherein the molecules or compounds are selected from DNA
molecules, RNA molecules, peptide nucleic acids, artificial chromosome constructions, peptides,
transcription factors, repressors, and regulatory molecules.
13. A purified protein or a portion thereof selected from:
 - a) an amino acid sequence of SEQ ID NO:1;
 - 35 b) an antigenic epitope of SEQ ID NO:1; and

WO 02/059311

PCT/US01/48917

- c) a biologically active portion of SEQ ID NO:1.
14. A composition comprising the protein of claim 13.
15. A method for using a protein to screen a plurality of molecules or compounds to identify at least one ligand, the method comprising:
- 5 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
- 10 molecules, RNA molecules, peptide nucleic acids, peptides, proteins, mimetics, agonists, antagonists, antibodies, immunoglobulins, inhibitors, and drugs.
17. A method of using a protein to prepare and purify antibodies comprising:
- a) immunizing a animal with the protein of claim 13 under conditions to elicit an antibody response;
- 15 b) isolating animal antibodies;
- c) attaching the protein to a substrate;
- d) contacting the substrate with isolated antibodies under conditions to allow specific binding to the protein;
- e) dissociating the antibodies from the protein, thereby obtaining purified antibodies.
- 20 18. An antibody produced by the method of claim 17.
19. An isolated protein comprising SEQ ID NO:1.
20. A purified antibody which binds specifically to the protein of claim 19.
21. A method for preparing and purifying a polyclonal antibody with the specificity of the antibody of claim 20 comprising:
- 25 a) immunizing a animal with a protein of SEQ ID NO:1 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
- 30 to the protein;
- e) dissociating the antibodies from the protein, thereby obtaining purified polyclonal antibodies.
22. A polyclonal antibody produced by the method of claim 21.
23. A method for preparing a monoclonal antibody with the specificity of the antibody of claim 20
- 35 comprising:

WO 02/059311

PCT/US01/48917

- a) immunizing a animal with a protein of SEQ ID NO:1 under conditions to elicit an antibody response;
- b) isolating antibody-producing cells from the animal;
- c) fusing the antibody-producing cells with immortalized cells in culture to form monoclonal antibody producing hybridoma cells;
- 5 d) culturing the hybridoma cells; and
- e) isolating monoclonal antibodies from culture.
24. A monoclonal antibody produced by the method of claim 23.
25. A method for using an antibody to immunopurify a protein comprising:
- 10 a) attaching an antibody of claim 20 to a substrate,
- b) exposing the antibody to a sample containing protein under conditions to allow antibody:protein complexes to form,
- c) dissociating the protein from the complex, and
- d) collecting the purified protein.
- 15 26. A method for using an antibody to detect expression of a protein in a sample, the method comprising:
- a) combining the antibody of claim 20 with a sample under conditions which allow the formation of antibody:protein complexes; and
- b) detecting complex formation, wherein complex formation indicates expression of the
- 20 protein in the sample.
27. The method of claim 2 wherein complex formation is compared with standards and is diagnostic of a colon disorder, particularly colon cancer, or colon polyps.
28. A composition comprising an antibody of claim 20 and a labeling moiety.
29. A composition comprising an antibody of claim 20 and a pharmaceutical agent.


```

441      450      459      468      477      486
TCC TGC CTT CCT AGG GTT GAG AGC GGC ACT GCC ATG GCT TCT CTG GAC GAC CCA
-----
495      504      513      522      531      540
GGG GAA GTG AGG GAG GGC TTC CTC TGC CCT CTG TGC CTG AAG GAT CTG CAG TCT
-----
G   E   V   R   E   G   F   L   C   P   L   C   L   K   D   L   Q   S
-----
549      558      567      576      585      594
TTC TAT CAG CTT CAC TCA CAT TAC GAG GAA GAA CAC TCA GGG GAA GAC CGT GAT
-----
F   Y   Q   L   H   S   H   Y   E   E   E   H   S   G   E   D   R   D
-----
603      612      621      630      639      648
GTC AAA GGG CAA ATT AAA AGT CTT GTC CAG AAG GCT AAA AAG GCA AAG GAC AGG
-----
V   K   G   Q   I   K   S   L   V   Q   K   A   K   K   A   K   D   R
-----
657      666      675      684      693      702
TTG TTG AAA CGA GAA GGG GAT GAT CGA GCA GAG TCA GGG ACC CAA GGA TAT GAG
-----
L   L   K   R   E   G   D   D   R   A   E   S   G   T   Q   G   Y   E

```

FIGURE 1B

```

711      720      729      738      747      756
TCT TTC AGC TAT GGA GGG GTT GAT CCT TAC ATG TGG GAA CCC CAG GAG CTT GGT
-----
S F S Y G G V D P Y M W E P Q E L G

765      774      783      792      801      810
GCT GTG AGA AGC CAT CTT TCC GAC TTC AAA AAA CAC CGA GCT GCT AGA ATT GAC
-----
A V R S H L S D F K K H R A A R I D

819      828      837      846      855      864
CAC TAT GTT GTG GAA GTC AAT AAA CTA ATA ATC AGG TTA GAG AAG CTC ACT GCA
-----
H Y V V E V N K L I I R L E K L T A

873      882      891      900      909      918
TTT GAC AGA ACA AAT ACT GAG TCT GCA AAG ATT CGA GCA ATA GAA AAG TCT GTG
-----
F D R T N T E S A K I R A I E K S V

927      936      945      954      963      972
GTG CCT TGG GTC AAC GAC CAG GAT GTC CCT TTC TGT CCA GAC TGT GGG AAT AAG
-----
V P W V N D Q D V P F C P D C G N K

```

FIGURE 1C

WO 02/059311

PCT/US01/48917

4/22

```

981 1035 1035 1044 1053 1062 1071 1080
TTC AGC ATC CGG AAC CGC CGC CAC CAC TGC CGC CTC TGC GGG TCT ATT ATG TGC
F S I R N R R H H C R L C G S I M C
1035 1044 1053 1062 1071 1080
AAG AAG TGT ATG GAG CTC ATC AGC CTT CCC TTG GCA AAC AAG CTC ACC AGT GCC
K K C M E L I S L P L A N K L T S A
1089 1098 1107 1116 1125 1134
AGC AAG GAG TCC CTG AGC ACC CAC ACC AGC CCC AGC CAG TCA CCC AAC AGT GTC
S K E S L S T H T S P S Q S P N S V
1143 1152 1161 1170 1179 1188
CAT GGC TCC CGC CGA GGC AGC ATC AGC AGC ATG AGC AGT GTC AGC TCG GTC CTG
H G S R R G S I S S M S S V S S V L
1197 1206 1215 1224 1233 1242
GAT GAG AAG GAC GAT GAC CGG ATC CGC TGC TGT ACA CAC TGC AAG GAC ACG CTG
D E K D D D R I R C C T H C K D T L

```

FIGURE 1D

```

1251      1260      1269      1278      1287      1296
CTC AAG AGA GAG CAG CAG AAT GAT GAG AAG GAG CAC ACA CCT GAC ATC GTG AAG
L K R E Q Q I D E K E H T P D I V K
-----
1305      1314      1323      1332      1341      1350
CTC TAC GAG AAA TTA CGA CTT TGC ATG GAG AAA GTT GAC CAG AAA GCT CCA GAA
L Y E K L R L C M E K V D Q K A P E
-----
1359      1368      1377      1386      1395      1404
TAC AUC AGG ATG GCA TCA TTA AAT GCT GGG GAG ACA ACC TAC AGT CTG GAA
Y I R M A A S L N A G E T T Y S L E
-----
1413      1422      1431      1440      1449      1458
CAT GCC AGT GAC CTT CGA GTG GAA GTG CAG AAA GTG TAT GAG CTG ATA GAC GCT
H A S D L R V E V Q K V Y E L I D A
-----
1467      1476      1485      1494      1503      1512
TTA AGT AAG AAG ATC TTA ACC TTG GGC TTG AAC CAG GAC CCT CCA CCA CAT CCA
L S K K I L T L G L N Q D P P P H P
-----

```

FIGURE 1E

WO 02/059311

PCT/US01/48917

6/22

```

1521      1530      1539      1548      1557      1566
AGC AAT TTG CCG CTG CAG AGA ATG ATC AGA TAC TCA GCT ACA CTT TTT GTG CAG
-----
S N L R L Q R M I R Y S A T L F V Q
-----
1575      1584      1593      1602      1611      1620
GAA AAG TTG CTT GGT TTG ATG TCA CTG CCA ACC AAA GAA CAG TTT GAG GAA CTG
-----
E K L L G L M S L P T K E Q F E E L
-----
1629      1638      1647      1656      1665      1674
AAA AAG AAA AGG AAG GAG GAA ATG GAG AGG AAG AGG GCC GTG GAG AGA CAA GCT
-----
K K K R K E E M E R K R A V E R Q A
-----
1683      1692      1701      1710      1719      1728
GCC CTG GAG TCC CAG CGA AGG CTT GAG GAA AGG CAG AGT GGC CTG GCT TCT CGA
-----
A L E S Q R R L E E R Q S G L A S R
-----
1737      1746      1755      1764      1773      1782
GCG GCC AAC GGG GAG GTG GCA TCT CTC CGC AGG GGC CCT GCC CCC TTG AGA AAG
-----
A A N G E V A S L R R G P A P L R K

```

FIGURE 1F

WO 02/059311

PCT/US01/48917

7/22

```

1791      1800      1809      1818      1827      1836
GCT GAG GGC TGG CTC CCA CTG TCA GGA GGT CAG GGG CAG AGT GAG GAC TCA GAC
-----
A E G W L P L S G G Q G Q S E D S D
-----
1845      1854      1863      1872      1881      1890
CCG CTC CTC CAG ATC CAC AAC ATC ACA TCA TTC ATC AGG CAG GCC AAG GCC
-----
P L L Q Q I H N I T S F I R Q A K A
-----
1899      1908      1917      1926      1935      1944
GCG GGC CGC ATG GAT GAA GTG CGC ACT CTG CAG GAG AAC CTG CGG CAG CTG CAG
-----
A G R M D E V R T L Q E N L R Q L Q
-----
1953      1962      1971      1980      1989      1998
GAC GAG TAT GAC CAG CAG ACA GAG AAG GCC ATC GAG CTG TCC CGG AGG CAG
-----
D E Y D Q Q Q T E K A I E L S R R Q
-----
2007      2016      2025      2034      2043      2052
GCT GAG GAG GAG GAC CTG CAG CGG GAA CAG CTG CAG ATG TTG CGT GAA CGG GAG
-----
A E E E D L Q R E Q L Q M L R E R E
-----

```

FIGURE 1G

WO 02/059311

PCT/US01/48917

8/22

```

2061      2070      2079      2088      2097      2106
TTG GAA CGA GAA AGG GAG CAG TTT CGG GTG GCA TCC CTG CAG ACA CGG ACT CGG
-----
L E R E R E Q F R V A S L H T R T R
-----
2115      2124      2133      2142      2151      2160
TCC CTG GAC TTC AGA GAA ATC GGC CCT TTT CAG CTG GAG CCC AGC AGA GAG CCT
-----
S L D F R E I G P F Q L E P S R E P
-----
2169      2178      2187      2196      2205      2214
CGC ACC CAC CTT GCT TAT GCT TTG GAT CTA GGC TCT TCC CGA GTT CCA AGC AGC
-----
R T H L A Y A L D L G S S P V P S S
-----
2223      2232      2241      2250      2259      2268
ACA GCT CCC AAG ACC CCT TCA CTT AGC TCA ACT CAA CCC ACC AGA GTG TGG TCT
-----
T A P K T P S L S S T Q P T R V W S
-----
2277      2286      2295      2304      2313      2322
GGG CCC CCA GCC GTT GGC CAG GAG CGC TTA CCC CAG AGC AGC ATG CCA CAG CAA
-----
G P P A V G Q E R L P Q S S M P Q Q

```

FIGURE 1H

WO 02/059311

PCT/US01/48917

9/22

```

2331      2340      2349      2358      2367      2376
CAT GAG GGG CCC TCC TTA AAC CCC TTT GAT GAG GAA GAC CTC TCC AGC CCC ATG
-----
H E G P S L N P F D E E D L S S P M

2385      2394      2403      2412      2421      2430
GAA GAG GCC ACT ACT GGT CCT CCT GCT GCA GGG GTT TCC TPA GAC CCT TCA GCC
-----
E E A T T G P P A A G V S L D P S A

2439      2448      2457      2466      2475      2484
CGC ATC CTG AAA GAG TAC AAT CCT TTC GAG GAA GAG GAC GAG GAG GAG GAA GCA
-----
R I L K E Y N P F E E E D E E E A

2493      2502      2511      2520      2529      2538
GTG GCA GGG AAT CCA TTC AAT CAG CCA GAC AGC CCA GCT CCT AAC CCC TTC AGT
-----
V A G N P F I Q P D S P A P N P F S

2547      2556      2565      2574      2583      2592
GAG GAA GAC GAA CAT CCC CAG CAG AGG CTC TCA AGC CCT CTG GTT CCT GGT AAC
-----
E E D E H P Q Q R L S S P L V P G N

```

FIGURE 11

WO 02/059311

PCT/US01/48917

10/22

```

2601      2610      2619      2628      2637      2646
CCC TTT GAG GAA CCC ACC TGT ATC AAC CCC TTT GAG ATG GAC AGT GAC AGT GGG
-----
P F E E P T C I N P F E M D S D S G
-----
2655      2664      2673      2682      2691      2700
CCA GAG GCT GAG GAG CCC ATA GAG GAA GAG CTC CTC CTG CAG CAG ATC GAT AAC
-----
P E A E E P I E E E L L L Q Q I D N
-----
2709      2718      2727      2736      2745      2754
ATC AAG CCA TAC ATC TTT GAT GCC AAG CAG TGC GGC CGC CTG GAT GAG GTA GAG
-----
I K A Y I F D A K Q C G R L D E V E
-----
2763      2772      2781      2790      2799      2808
GTG CTG ACA GAG AAT CTG CGG GAG CTG AAG CAC ACC CTG GCC AAG CAG AAG GGG
-----
V L T E N L R E L K H T L A K Q K G
-----
2817      2826      2835      2844      2853      2862
GGC ACT GAC TGA CCA GCA GTG GAG AGG GCA CCT TTG GGC CCA GGG GTC TGG CAG
-----
G T D

```

FIGURE 1J

WO 02/059311

PCT/US01/48917

11/22

2871 2880 2889 2898 2907 2916
 GAG CCA GTG GAG CAG GAC AGA GGG CAG GCA GGA TGG ATG GGG AAG GTG GCA GGG
 2925 2934 2943 2952 2961 2970
 TGA GAA CTC AGA TGC ACA CAG GTG AGG GGC AGG AAT CTG CTG TTT TGT GTT GCG
 2979 2988 2997 3006 3015 3024
 CAC TTT GAG GTA TTT CCA CTA CAG TTG AAT AAT AAA ATA GAA ACT AGA ACA GGG
 3033 3042 3051 3060 3069 3078
 AGA ATC AGC ATT CAG TTG CTG CTT TTC CTG TTT ATT ATT ACT ATC TTT TGT AAT
 3087 3096 3105 3114 3123 3132
 CGG AGG TTT ACC CCT TTT GAA GGG ACT TTA CAT TTT TAC TAC CGA GAT ATA ACT
 3141 3150 3159 3168 3177 3186
 AAA TGC AGC TCT GTT GGG CCC AGG GCA GAA ATG GCT GCT CTG TAC CTC TTG GGT
 3195 3204 3213 3222 3231 3240
 CCA TTT GCT ACT GCC TAG TCT TGG TTC CTT ATG CAG TAT TAT AGG GCA GCC TTT
 3249 3258 3267 3276 3285 3294
 TTA GAG CCC TTC CTT TAG CCA AGA CAG AGA TAG ATT CCA CTG AGC TCT ATT

FIGURE 1K

WO 02/059311

PCT/US01/48917

12/22

3303 3312 3321 3330 3339 3348
 CTG CTC TGA CAG AAG TCC ATC CCT AGT AGG CTG TGA GTT CCA TTT CAC CTG GGG
 3357 3366 3375 3384 3393 3402
 CCG CCT CTC CCC TGC TCT GCA CTT CCT GTC TGT ACA ATA GAA GGG GGA GGT GGT
 3411 3420 3429 3438 3447 3456
 GCT ATG AAG GGG AGA GTT TAG ACC CAG GAG AGC CCA GCA CCT CTC TTT AAG GTG
 3465 3474 3483 3492 3501 3510
 GGG TGA TGG GAA TAT TTC ACC AGG GTC TAT TTT CTC AGT TTA AGT TCT TTT TTG
 3519 3528 3537 3546 3555 3564
 TCT CTT TCA GGA AGT TAA GCT CCC AGT GCA GGG TAT CAA TGT GAA TCT GGT CCT
 3573 3582 3591 3600 3609 3618
 GAG CTT TTT AGA AAA TAA GAG TGG TGG CCG GGC GCG CTG GCT CAC GCC TGT AAT
 3627 3636 3645 3654 3663 3672
 CGC AAC ACT TTG GGA GGC CGA GGC AGG CAG ATC ACA AGG TCA GGA GAT CGA GAC
 3681 3690 3699 3708 3717 3726
 CAC CCT GGC TAA CAC AGT GAA ATC CCG TCT CTA AAA ATA CAA AAA AIT AGC

FIGURE 1L

WO 02/059311

PCT/US01/48917

13/22

3735 3744 3753 3762 3771 3780
 CAG GCG TGG TGG TGG CCT GTG GTC CCG GCT ACT TTG GAG GCT GAG GCA GGA
 3789 3798 3807 3816 3825 3834
 GAA TGC TGT GAA CCC GGG AGG CAG AGC TTG CAG TGG GCC GAG ATC GCA CCA CTG
 3843 3852 3861 3870 3879 3888
 CAC TCC AGC CTG GGC AGT GAG AGT GAG ACT CCG TCT CAA AAA AAA ATA AGA GTG
 3897 3906 3915 3924 3933 3942
 GTT AAC TTT TTG GTT GAT AAT AGT AAT AAT TGT GAG AAG AGT TTC CCC ACC CTT
 3951 3960 3969 3978 3987 3996
 CTT CTG ATA GAA CAA TTG TCT GTC ACT GGA GAA ATC TCC CTC CAG AGC TTT GGC
 4005 4014 4023 4032 4041 4050
 AAA GTT ACT CTG ATC TGG GTC TGT TTA AAA GGC CGG GTC TCT AAT TTA GGA ATC
 4059 4068 4077 4086 4095 4104
 GGT GAT TTG GAA GCT TTT GCA GAA CAT CAC CAG AAG AAG GGA AGC TTC CCA GAG
 4113 4122 4131 4140 4149 4158
 TCA GAA GCT AAA TTA AAA TAA TTT CCA AGG CTA TTT GAT CAG CCT TCT TCC TTT

FIGURE 1M

WO 02/059311

PCT/US01/48917

14/22

4167 4176 4185 4194 4203 4212
 TTG GTT CAT TGT GTC CTG ACT TGG GGC ACT GAT GAG ATT TTT TAT TTT TTT TGA
 4221 4230 4239 4248 4257 4266
 GAC GGA GTT TCA CTT TGT TGC CCA GGC TGG AGT GTA GTG GCA TGA TCT CGG CTC
 4275 4284 4293 4302 4311 4320
 ACT GCA ACC TCC ACC TCC CTG GTT CAA GTG ATA TTT GTG CCT CAG CCC CCT GAA
 4329 4338 4347 4356 4365 4374
 TAG CTG GGA TTA CAG GCG TGG GCC ACC ACG CCT GGT TAA TCT CTG TAT TTT TGG
 4383 4392 4401 4410 4419 4428
 TAG AGG TAG AGT TTC ACC ATT TTG GCC AGG CTG GTT TCG AAC TCC TGG CCT CAA
 4437 4446 4455 4464 4473 4482
 GTG ATC TGC CAC CTT GGC CTC CCA AAG TGC TGG GAA TAC AGG CGT GAG CCA CTG
 4491 4500 4509 4518 4527 4536
 TGC CCG CCC TGA TGA GAT ATT TTA TTA CCA ATG TTA GTA TTG AGA AAC TGA AAT
 4545 4554 4563 4572 4581 4590
 GTT TGA AGA AGC ACA ACC CAG GAT CGT GCT GGT AGC ACC ACA GTA CTT AAA CTG

FIGURE IN

WO 02/059311

PCT/US01/48917

15/22

4599 4608 4617 4626 4635 4644
 TTG GTC AAT TAA GGC CAG AAA GGG AAA TTG TTA ATT TAG CTC TGG TGC TTT GGT
 4653 4662 4671 4680 4689 4698
 TTA CAG GAA CAT AAC TCT TAA CTG ACA TCT GAC ATC ATG ATA GCC ATA TGT GCT
 4707 4716 4725 4734 4743 4752
 CAG CTC TGG GTA GAG TTT CTG CAG TTA CTC ATC TGA ACT AAT GAA CAA TAA CTG
 4761 4770 4779 4788 4797 4806
 ACC ACT AGT CAC TTT ATG CCG TGT AAC TAG CTC TAG GCC ATA CTT TCA CTG GTT
 4815 4824 4833 4842 4851 4860
 ACT GGT GTC AAA GCT GAA AAT CAT TTT GTT ACA TTC TGG TGA AGA CCC CTC TTG
 4869 4878 4887 4896 4905 4914
 ATA ATG GGA ATG TTT TAA CTC TCT TGA TGA AAA AAT AAT CIG TAT TTG TGT TGA
 4923 4932 4941 4950 4959 4968
 TGT TCA CAT TTC TGT AGC ACA TTT CTT ATC CTT TTG GTT GAA TGA AAA GAT CTT
 4977 4986 4995 5004 5013 5022
 GTA TAG GGG TGT GGA GAT GGG GAG TGG GTA GAA GTG TGT GAA CGC TTT GCA

FIGURE 10

WO 02/059311

PCT/US01/48917

16/22

5031 5040 5049 5058 5067 5076
TTT GGG ATC TGT TCA CAA ACA GCC ATA TGA GTG TGT TAA TGA ATG TCA GCC AGT
5085 5094 5103 5112 5121 5130
TAC CAA CCC TGC TGG TTG TTA TGG GTT GTT TTG AGA AGT TGG CAA CCA GCC ATC
5139 5148 5157 5166 5175 5184
TAA GAT GTT GCC TGG TAC AGG CCT CTT TTC CCT GAG GCC CAT GAC ATT TCT
5193 5202 5211 5220 5229 5238
CTG TTA CTC TAG AGG GTT TCT CAG ATG GCC AGT AGG CTC CTC CCT GCT TAG GGT
5247 5256 5265 5274 5283 5292
CTC ATT TCT CTG AAA AGA GGA TGA ACT GAA AAG CAG GTA GTT CCA GAA GCT AAT
5301 5310 5319 5328 5337 5346
TGC TGC TTT CCA TCA TAA TTA TTT TTC TTG TGA GAA CAT TTC TCT TTT AAT TAG
5355 5364 5373 5382 5391 5400
CTA GTG ATT TTG ATT AAG ACT AAT TCA CTA AAC ATA CCC TTC CCT CAA ATC ACC
5409 5418 5427 5436 5445 5454
TCA GGT AGC AAT CTG TAA GTA ACT AAA AGC ATT GAA AAA CAC ACA AGA AAC ATT

FIGURE 1P

WO 02/059311

PCT/US01/48917

17/22

5463 5472 5481 5490 5499 5508
TTT AAC TAT TTT TAA AGG CCT GGC CGG GTG CAG TGG CTC ACT CCT GTA ATC
5517 5526 5535 5544 5553 5562
CCA GCA CTT TGG GAG GCC GAG ACG GGT GGA TCA CCT GAG GTC AGG AAT TTG AGA
5571 5580 5589 5598 5607 5616
CTA GCA TGG CCA ACA TGG TGA AAC CCC GTG TCT ACT AAA AAT ACA AAA AAT AGC
5625 5634 5643 5652 5661 5670
TGG GCG TGG TGG CAG ACG CCT GTA ATC CCA GCT ACT CAG GGA GGC TGA GGC ACG
5679 5688 5697 5706 5715 5724
AGA AUC GCT TGA ACC TGG GAG GCG GAG GTT GCA GTG AGC CAA GAT TGC GCC ACT
5733 5742 5751 5760 5769 5778
GCA CTC CAA CCT GGG TGA CAA GAG CAA GAC TCC GTC TCA CAA AAA ACA AAC AAA
5787 5796 5805 5814 5823 5832
ACT GAC TGA TTG AAT ATA CTA AAC CAA ACT AAA ATC ATA TTC TTT TGA TTA AGT
5841 5850 5859 5868 5877 5886
TTA TCC ATG GCT GTA TTC TTC TAT GAA TTC TTC CAT TTA AAT CTT CCA CGA TTA

FIGURE 1Q

WO 02/059311

PCT/US01/48917

18/22

5895 5904 5913 5922 5931 5940
TCT TCT CCT GTA AAT ACA TCA CAG GAG TTA GAA TTC TCT ACC CAT CAG CTG TAC

5949 5958 5967 5976 5985 5994
CAT GTC GCA GAA TTC ATG CAG GCA CAA AGT TGG AGT TAC AGA GAT GGG TTG ACA

6003 6012 6021 6030 6039 6048
GCA GGC AAA CTT GGC CTA TGT AAT ATA ACC ACA ACT TCA AGT TCT TAC CTC ATG

6057 6066 6075 6084 6093 6102
TGA ATA TTC ACC CTT TCT TTA GTC TTC CAA GGC AAA CAG CCC CGT CTC ATC ACC

6111 6120 6129 6138 6147 6156
AGA TGA GCA AGG TCT TGA TAT GGC ATA GCA GAT CTC CCT AGA CAC AGA TCA TGA

6165 6174 6183 6192 6201 6210
GAA AAG ATG GAA GAG ACT TAG GGA TTC AGG CAT CAG ATG AAG TTG GCT TTT CCC

6219 6228 6237 6246 6255 6264
TTT TAT GCC TTG TTT GTA TTT ACC CTG TCT AAT ACA CTA AGG ATA CTT ACT CAT

6273 6282 6291 6300 6309 6318
TGT ACT TGC AGC TCA ATA TGT CTT TGC TGT TCA GAT ACT AAA ATG TAC CTC TGA

FIGURE 1R

WO 02/059311

PCT/US01/48917

19/22

6327 GTC ATT GTG AGC TGT GTG GTA GGT TGG 6345 6354 6363 6372
 AAA ATG AAA AGG TGG TCT CTT TAC CAG GTC ACA GAC TGT AGC AGA TTG TGC TTG 6426
 TTA TCT GAC AAT GAC TGT CAC TTT GAG GGT CGT TGA TTT GCA TGC ACT ACT CTG 6480
 GGG CCT TGT ATT GGA GCC TTT TTT AAA AAA AAT AAA ATC TGA GAT AGA GGT TGG 6534
 GGG TGT CTG TGT TGT GCA CAT GTG TTA CAA GTG AGA ATC ATC AGA TGA CAT 6588
 CCC TTC TCC TTT CTT GAT GAC AAC CAT CTG AGT ATC AGA ATA GTT CCA GCA CCT 6642
 GTG TTG TTT GTC TGG TTA AGG CCT CTG GAA AAA ATG AAG GTC ACT GGG TTC TGA 6696
 ACA GGG GAT AGA TAC GGG TTC CAG TTC TGC CCT TAT TCC CAG TTA TTC CTG CAG 6750

FIGURE 1S

WO 02/059311

PCT/US01/48917

20/22

6759 TGC TGG TTA AAT GAA CAG TTT TTAC AGA AGT ATA GTG CAG ATC CCT TAA TGA TCC 6804
6813 TAT CAT CCC AGT TTC ACA CTG GAA GAA ACT GAG GCT CAG AGA TTG AGG AAC CTT 6858
6867 GTG ACT GCC ATC TGT GCC AGT CAC CAC CTC TGC ATG ACC ACA TTT CCC TCC ATT 6912
6921 AGC ACT ACC AGC ATG CCT GTG AGG TAG TTA CTC AGC TGT TCA TCT GAC CCC AGA 6966
6975 CGT AGA AAA GTT AAG AGA GAT TAG CCT GGC TTA GGC CAC ACA GCC AGT CAG TGG 7020
7029 TGG AAT GAG GGT TTA AAT CCA GAT CTG CCT GAC TTA ATT TGC CCT CAG TAA TAA 7074
7083 GTA GCT GGG TCA GGT AGT GGT GTG GTG GTG GGT GCT CAA CCA TTG TGT CAA 7128
7137 GCT CAA AGA GGG CAG AAA AGG GCA ATG GGG AAT GAG GTT GGG TGC TGC TGA GCT 7182

FIGURE 1T

WO 02/059311

PCT/US01/48917

22/22

7623 7632 7641 7650 7659 7668
 CAG ACA CGT AGA GTT TGA AAG TTA TCC TAT CCT TAC CAC AAG AAA AAG CTG GCC

 7677 7686 7695 7704 7713 7722
 AAA CTC AAA AGC ATT GGC CTT TCT TGG TCC CAT TAG AGA GCT GAG TGT GCA GGT

 7731 7740 7749 7758 7767 7776
 CAA ACT GCC ATC CCA AAA TCT AGA GAG AGG TGA GTC CAG AGA GTG AGC AAG ATC

 7785 7794 7803 7812 7821 7830
 TAA TTG CCT GGA GCA GAA GCC ACG GGA GCC ATA GCT AGT AGG AAA ACT TAA ATG

 7839 7848 7857 7866 7875 7884
 GTA ATT TTG ATG AAC TGC TGG AGG TTG AGT CTG GAC TAG CAT GGG AGG GAG CAA

 7893 7902 7911 7920 7929 7938
 CTC CCC CTC ACC CCC AAC GCT TTC CTG GGT TTT ACC CTA CTT GGT TCT ATT GGT

 7947 7956 7965 7974 7983 7992
 GAA GAG CTG AGG AAG ACC CTC TGG TGA CTC TGG CAG GGG AAA GGG AGA ATC ATC

 8001 8010 8019
 TTT GTA ATC AGA GCC TTC TCC ATA ATA AA 3'

FIGURE IV

WO 02/059311

1/18

PCT/US01/48917

<110> INCYTE GENOMICS, INC.
NOLVEN, Amy
KRASNOW, Randi E.
WARREN, Bridget
BAUGHN, Mariah R.

<120> COLON CANCER MARKER

<130> PC-0031 PCT

<140> To Be Assigned
<141> Herewith

<150> 09/740,235
<151> 2000-12-18

<160> 36

<170> PERL Program

<210> 1
<211> 784
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5098390CD1

```

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

```


WO 02/059311

3/18

PCT/US01/48917

Glu Met Asp Ser Asp Ser Gly Pro Glu Ala Glu Glu Pro Ile Glu	725	730	735
Glu Glu Leu Leu Leu Gln Gln Ile Asp Asn Ile Lys Ala Tyr Ile	740	745	750
Pro Asp Ala Lys Gln Cys Gly Arg Leu Asp Glu Val Glu Val Leu	755	760	765
Thr Glu Asn Leu Arg Glu Leu Lys His Thr Leu Ala Lys Gln Lys	770	775	780
Gly Gly Thr Asp			

<210> 2
 <211> 8021
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 5098390Cb1

<400> 2
 gccctcccgag ttccgagcga ccgatggaga tggcggctgc ggctgagtga cggacggtgg 60
 aggccccagag ccggggcctg aaggggggga caaacctggg tgcccgcagc agccccggcag 120
 gglgtctttac aagtatcaag aacttactat atgtggttga ataaacaatc asggtaaga 180
 gcatacaagta aaaactctctg ctgttgtata agtacttcag acatccccc agtggctgaa 240
 gtggcatatg aattatgaag ttgatcatt tggatgaat gtaagagaat tggcaaggc 300
 tccctctact ccagagagga aacctcatcc agggccatga agccactcc tcaccactctg 360
 tggctgctt aagctaatgc tgcgggaacc atggttccct gggagagatc aagctgactc 420
 ttggcatgag atctctgctt hcttaggtt gagagcgga ctggcatggc ttctctggac 480
 gaccacaggg aagttaggga ggcttctctg tgcctctctg gctgaagga tctgactct 540
 ttctatcagc ttactcaca ttacagagaa gaacctcag gggaaagacc tgatgtcaa 600
 ggccaatata aaagtcttct ccagaagct aaaaaagaa aggcaggtt gttgaaaca 660
 gaagggatg atcgagcaga tccagggacc caaggatag agtcttctag ctatggagg 720
 gttgatcctt acatgtgga acccaggag ctgtgtgctg tgaagaacca tottccogac 780
 ttcaaaaaac accagctgc tagaattgac cactatgttg tgaagtcaa taaactaata 840
 atcaggttag agaagctcac tgcattgac agaacaata ctgactgac aagattcga 900
 gcaatagaaa agtctgtggt gcttgggtc aacgaccagg atgtccctt ctgtccagc 960
 tctgggaata agttcagcat ccggaaccgc ccccaccact gccctctctg cgggtctatt 1020
 atgtgcaaga agttgatgga gctcatcagc ctctccctgg caaacaagct caccagtgcc 1080
 agcaaggag ccttgagcgc ccacaccagc ccagccagc caccacaag tctccatggc 1140
 tcccgccgag gacgatcag cagatgagc agtctcagct cgtctctgga tggagagac 1200
 gatgaccgga tccgctgctg tacactctgc aaggacagc tgcctcaaga agggcagcag 1260
 atgatgaga agggcacac acctgacatc gtgaagctct acgagaat acgactttgc 1320
 atggagaag ttgaccagaa agtccagaa tacatcagga tggcagcact attaaatgt 1380
 ggggagacaa cctacagctt ggaacatgcc agtgcacttc gagtggagc gcgaaagtg 1440
 tatgagctga tagacgcttt aagtaagaag atcttaacct tgggcttgaa ccaggaacct 1500
 ccaccacatc caagcaattt gggctgcaag agaatgatca gatactcagc taacttttt 1560
 gtgcagaaaa agttgcttgg ttgatgtca ctgccaacca aagaacagtt tgaggaaactg 1620
 aaaaagaaaa gaaggagga aatggagag aagggggcgg tggagagaa agctccctg 1680
 gagtcccaag gaaggcttga ggaaggcag agtggcctgg ctctctgagc ggccaacggg 1740
 gaagtgagca ctctccgag gggccctgcc ccttgagaa aggtgaggg ctgctccca 1800
 ctgtcagag gtcaggggca ggtgagagac tcagaccgc tccctcagca gatccacaac 1860
 atcaatcat tcaatcagca ggcaaggcc gggggccga tggatgaagt gggactctg 1920
 caggagaacc tggggagct gcaggacag tatgaccagc agcacacaga gaaggccatc 1980
 gactgtctcc ggaggcagc tgaggaggag gacctgcagc gggaaagct gcagatgtg 2040
 cgtgaacggg agttggaagc agaaagagc cagtctcagg tggactccct gcacacagc 2100
 actcagctcc tggactcagc agaatccgc ccttttcagc tggagcccag cagagagcct 2160
 cgcaccacc ttgcttatg tttggatcta ggtctctccc cagtccaag cagcacagct 2220
 cccaagaccc ctctacttag ctcaactcaa ccaccagag tglggtctgg gccccagcc 2280
 gtggccagc agcgttacc ccagagcagc atgcccagc aactgaggg gccctccta 2340
 aaccctttg atgaggaaga cctctccagc cccatggaag aggcactac tggctcctc 2400
 gctgcagggg ttctctaga cctctcagc cgcactctga aagagtcaa tcttctcag 2460
 gaagagcagc aggagggaga agcagtggca gggaaatccat tcaatcagcc agacagcca 2520

WO 02/059311

4/18

PCT/US01/48917

```

gctcctaacc ccttcagtga ggaagacgaa catccaccgc agaggctctc aagccctctg 2580
gttccttgta accccttga ggaacctacc tgttcaacc ccttlgagat ggcagtgac 2640
agtgggccag agctgagga gcccatagag gaagagctcc tcoctgacga gatccataac 2700
atcaagcaat acatcttga tgcacagcag tggggccccc tggatgaggt agagtgctg 2760
acagagaaat tggggagct gaagcacacc ctggccaagc aqaagggggg cactgactga 2820
ccagcaatgg agagggcacc ttggggcccc ggggtctggc agggaccagt ggcagcagca 2880
agagggcaag caggatggat ggggaaggtg gcaaggtgag aactcagatg cacacagtg 2940
aggggcagga atctgctgt ttgtgttggc cactttgagg tatttccact acagttgaal 3000
aataaaatag aaactagaac agggagaatc agcattcagt tgcgtgtttt octgtttatt 3060
attaactatc ttgttaatg gaggtttacc cttttgaaag ggaacttaca tttttactac 3120
cgagataaa ctaaatcgag ctctgttggg cccagggcag aaatggctgc tgtgtacctc 3180
ttgggtccat ttgtactctg ctactcttgg ttccctatgc agtattatag ggcagocctt 3240
ttagagccct tcccttagcc aagacagaga agatagatcc cactgagctc tattctgctc 3300
tgacagaagt ccatccctag taggctgtga gtccatttc acctggggcc gccctcccc 3360
tgcctgcacc ttccctgtg tacaatagaa gggggaggtg ctgctatgaa ggggagagt 3420
tagaccagag agagccagc acctctctt aagtggggt gatgggata ttccaccag 3480
gtctatlttt ttagtttaag ttctttttg tctctttcag gaagtttag tccagtgca 3540
gggtatcaat gtgaatctgg tcoctagctt tttagaaaat aagagtggg gccgggcgcg 3600
ctggctcaag cctgtaatg caacactttg gggggccag gcaagcagat cacaggtca 3660
ggagatcgag accaccctgg ctaaccagt gaaatccgt ctctactaaa aatacaaaa 3720
attagccagc cgtgtggtg ggtgctgtg gtcccggcta ctttggaggc tgagcgaga 3780
gaatcgtgtg aaccgggag gcagagctg cagtggggc agatgcacc actgcactcc 3840
agcctgggca gtgagagtg gaactcctct caaaaaaaa taagagtggt taacttttg 3900
gtgataata gtattatttg tgagaagagt tccccacc ttctctgat agacaattg 3960
tcgtactg gagaatctc ctccagagc ttggcacaag ttaactgat ctgggtctg 4020
ttaaaggcc ggtctctaa tttagaaac ggtgatttg aagcctttg agacatca 4080
cagaagaag gaagctccc agagtcaaaa gctaaatba aataatttc aaggctatt 4140
gactcagctt ctcccttltt ggtlcatgt gctcagctt ggggcactga tgagatttt 4200
tattttttt gagaagagt ttaactttg tgcacagct ggaatgact ggcataact 4260
cgctcaact caactccc ctccctggt caagtatct ttgtgctca gcccttga 4320
tagctggat tcaagcgtg ggcacacag cctgttaa ctctgtatt ttgttagag 4380
tagatgttca ccatltggc caggctggt tcaaatcct ggcctcaat gatctgccac 4440
cttggcctcc caaagtctg ggaatacag cgtgagccac tgtcccgc ctgatgat 4500
atlttattac caatgttag attgaaaac tgaatgttt gaagaagcacc aaccaggat 4560
cgtctgtgta gcaccacgt acttaaacgt ttggtcaat agggccagaa agggaattg 4620
ttaatltagc tctggctct ttgtttacag gaacataact cttaactgac atctgacatc 4680
atgatagcca tatgtgctca gctctgggta gatttctgc agtactcat ctgaactaat 4740
gaacaataac tgaccactag tcaacttatg cgtgttaact agctctagcc catacttca 4800
ctggttactg gttgaaaag tgaattcat ttgttaact tctggtgag acccctctg 4860
ataatgggaa tgttttaact ctctgtatga aaaaataac tgtatttgg ttgatgtca 4920
cattctgta gcaacttct tctcttltg gttgaatgaa aagatctgt atggggtgt 4980
ggagatggg agtggttag agtgtgtgaa ggaagcttgc cacttggat ctatcaaa 5040
acagcaat at ggtgtgtta atgaagtca gccagtacc aacctgctg gttgtatg 5100
gtgttltga gaagtggca accagctc taagatgttg cctgtcaag gccctttt 5160
ttccctgag cccatgact ttctctgtta ctctagagg ttctcagat ggcagtagg 5220
ctctccctg cttaggtct cacttctctg aaaagagat gaactgaaa gcaggtagt 5280
ccagaagcta atgtgctt tccatcataa ttatlttct tgtgagaaca ttctcttt 5340
aattagctag tgattttgat taagactaat tcaataaaca tacccttcc tcaaatcacc 5400
tcagtagca atctgtaagt aactaaaagc attgaaaacc acacaagaaa catttttaa 5460
aactatttt aaaggcctgg ccgggtgag tggctcact ctgtaatcc agcacttgg 5520
gagggcaga cgggtggatc acctgaggtc aggaatttga gactagcatg gccacaatg 5580
tgaaccctcg tgttactaa aaatacaaaa attagctgg cgtgggtgca gacgctgta 5640
atcccagcta ctccggggg ctgagggcagc agaactcgtt gaacctggga ggcggaggt 5700
gcagtgagc agatltgag caclgacac caaactgggt gacaagaga agcctcgtc 5760
tcaaaaaaa caaacaaaac tgaactgatt aatatctaa accaaactaa aatcaatc 5820
tttggataa gttatccat ggtgtatcc ttctatgat tcttccatt aattttca 5880
cgattatctt ctccgtaaa tacaacacag gatttagaatt tcttaocca tcaagtgac 5940
catgctgag aattctgca gccacaagt tggagtaca gagatgggt gacagcaggc 6000
aaacttggcc tatgtattat aaccacaact tcaagttott acctcatgt aatattcacc 6060
cttcttttag tottccaaag caaacagccc cgtctcaaca ccagatgagc aaggtcttga 6120
tatggcatag cagatctccc tagacaacaga tcaatgaaaa agatggaaag gacttagga 6180
tcaagcatic agatgaagt gcttltccc ttttatgctc ttltgtatt taccctgct 6240
aataactaa ggatactac tcaattgact tgcagctcaa fatgtcttg ctgtcaagat 6300

```

WO 02/059311

5/18

PCT/US01/48917

```

actaaaaatgt acctctgagt cattgtgagc tgtgtggtag gttggacatt ggcatagttg 6360
gtcatggggac tcaaaaatgaa aaggtggctc cttaccagag tccagagctg tagcagattg 6420
tgcctgttat ctgcaaatga ctgtccactt gagggtccctt gatttgcctg caactactctg 6480
ggggcttcta ttggagccctt ttttaaaaaa aataaaatctt gagatagagg ttgggggtgt 6540
gtgtgtctgt gtgcacatgt gttaacaagt agaatactca gatgacatcc ctctctcttt 6600
ctrtgatgata accatctgag tatcagaata gttccagcac ctgtgttgtt tgtctgggta 6660
aggcctctgg aaaaaatgaa ggtcaactgg ttctgaacag gggatagata ogggttccag 6720
ttctgcccctt attcccagtt attcctgcaag tgcctgttaa atgaacagtt ttacagaagt 6780
atagtgcaaga tcccctaatg atcctatcat cccagtttca cactggaaga aactgaggct 6840
cagagattga gaaaccttgt gactgccatc tgtgccagtc accacctctg catgaccaca 6900
tttccctcca ttagcactac cagcaatgct gtgaggtagt taactcagct ttcatctgac 6960
cccagacgta gaaaagttaa gagagattag cctggcttag gccacacagc cagtcaatgg 7020
tggaaatgag gtttaaatcc agatctgect gacttaattt gccctcagta ataagtact 7080
gggtcaggta gtgggtggtt ggtgtgggtt gctcaacctt tgtgtcaagc tcaaaagagg 7140
cagaaaaagg caatggggaa tgaggttggg tgcctgctgag ctggtttccg atagtactg 7200
ccctgagctg tgaactgag acctccctag ggccttaaga atataatag tcttaaaaca 7260
ggagagtata atattctgtg ctactgtag gctagctga gtgtaagta ttacgcatgg 7320
cttactctag tcttcccaac aacctatba cacatgcaga aatgagttag tggtagatt 7380
aaactgtctt agggttgac tgcctgggtt cccagctagg tctgtctgac tgcataaac 7440
aagctgcctt agattgtgtt gtatgtggac accatgttt ctaatgtgga atgggtgtt 7500
tcatccctag caagtgcctc ctgtggaatt ttaagatgaa aaacttctgc ctaactctc 7560
aggaagccta gcgtagtgtt aaaaatactt cagaagcac ttctgggttt agctcagaca 7620
cgtagagtgt gaaagtatc ctatccttac cacaaagaaa agctgggcaa actcaaaa 7680
attgcctctt ctgtgtccca ttagagagct gagtgtgca gtcaactgc catcccaaaa 7740
tctagagaga ggtgagtcca gagagtgagc aagatctaat tgcctggagc agaagccag 7800
ggagccatag ctagtggaa aacttaaatg gtaattttga tgaactgctg gaggttgagt 7860
gtggactagc atggggagga gcaactcccc ctacccecca acgcttctct gggttttacc 7920
ctacttggtt ctattgtgga agagctgagg aagacctctt ggtgactctg gcaggggaaa 7980
ggggaatca tcttttaatt cagagccttc tccataataa a 8021

```

```

<210> 3
<211> 432
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 509839076

```

```

<400> 3
gccagagcc tcgcaaccac cttgcttatg ctttggatct aggotcttcc ccagttccaa 60
gccgacagc tcccagaacc ccttcaatta gctcaactca acccaacaga gtgtggtctg 120
ggccccagc cgtttggcac gagcctctac cccagagcac catgccaac caactatgag 180
ggcctctctt aaacctcttt gatgaggaag acctctccag ccccatggaa gaggccacta 240
ctgctctctc tgcctgaggg gtttctctag acctctcagc ccgcatcctg aagagatgaa 300
atcctttcga ggaagaggac gaagaggagg aagcagtggc agggaatcca ttcatcagc 360
cagacagccc agctcctaac ccttctagtg agggagagac aacatcccca gcagaggtct 420
tcaagcctc tg 432

```

```

<210> 4
<211> 573
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 8022266J1

```

```

<400> 4
gcctccgag ttccgagcga cagatggaga tggcggctgc ggctgagtga cggacgggtg 60
aggcccagag ccgggctctg aaggggggga caaacctggg tgcctgcagc agcccggcag 120
gggtctttac aagtatcaag aacttactat atgtggttga ataacaatc aaggtaaaga 180
gcatacaagta aaaaactctg cttgttgata agtactctag acattccccc astggctgaa 240

```

WO 02/059311

6/18

PCT/US01/48917

```

gtggcatalg aallatgaag ttggatcaatt tggatgaat ggtctctctt actccagaga 300
ggaaacctca tccagggcca tgaagccaat tctccacct ctgtgtgctg ctttaagctaa 350
tgcctgcggga acaatgggttc ctbtggagga atcaagctga ctcttggcat gagattctgt 420
ccttctcLaag gttgagagcg gcaactcccat ggtctctctg gaccaccCaq gggagtgag 480
ggagggttcc ctctgcccctc tgtgctgaa ggatctgcag tctttctate agottcaactc 540
acattaagag gaagaacctc cagggaaga cgg 573

```

```

<210> 5
<211> 598
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7394066H1

```

```

<400> 5
actgctcaatg ctgctgatgc tgcctcggcg ggagccatgg acactgttgg gtgactggct 60
ggggtctgtg tgggtgctca gggactcctt gctggcactg gtgagcttgt ttgccaaggg 120
aaggctgatg agctccatac actctcttga cataatagac ccgcagagcc ggcagtgtgt 180
cggcgggttc cggatgctga acttattccc acagtctgga cagaaggga catcctggtc 240
gttgacccaa ggcaccacag actttctat tgcctgaatc ttgcaagact cagtatttgt 300
tctgcaaat ccaagtgcct tctctaacct gattattagt ttattgactt ccacaacata 360
gtggtcaatt cttagcagctc ggtgtttttt gaagtggaa agatggcttc tcacagcacc 420
aagctcctgg ggttcccaca tgaaggatc aaacctcca tagctgaaag actcatalcc 480
ttgggtccct gactctgctc gatcaccctc tctctgttcc aacaactgtc ctttggcttt 540
ttagctcttc ggacaagact ttttaattgc ctttagatc acggctctcc cctgagtg 598

```

```

<210> 6
<211> 595
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 6309461H1

```

```

<400> 6
atggctctgt tagatcttca ctgcacctca tgaagaacctg cccagggaca aacagtgtgt 60
tgagaagcda tctttccgac ttcaaaaaac actgagctgc tagaattgac cactatgttg 120
tggaaagcda taaactaata atcaggttag agaatgcaac tgcatttgac agaacaaata 180
ctgggtctgc aaagattcga gcaatgaaa agctgtgtgt gcttggctc aacgaccagg 240
atgtcccttt ctgtccagac tgtgggaata agtccagcat ccggaaccgc cgcaccact 300
cgccctctg cgggtctatt atgtgcaaga agtqtatgga gctcatcagc cttcccttgg 360
caacaagct caccagtgc agcaaggagt ccttgagcac ccacaccagc cccagccagt 420
caaccaacag tgtccatggc tcccgcgag gcaagcatcag cagcatgagc agtctcagct 480
cgtccttga tgagaaggac gatgaccgga tccgctctg tacacactgc aaggacagc 540
tgcacaagag agagcagccg attgatgaga aggagccaca cctgaacatc gtgaa 595

```

```

<210> 7
<211> 645
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7436032H1

```

```

<220>
<221> unsure
<222> 20
<223> a, t, c, g, or other

```

WO 02/059311

7/18

PCT/US01/48917

```

<400> 7
tggcaactcata taggggccan tggggtcaact acatgcatgc tggagcggcc gccagttgtgc 60
tggaaagctct gtggtgccct gsgtcaacga ccagatgttc cctttctgtc cagaactgttg 120
gaabaagtctc agcatccgga accgcgccca ccactgcgcc ctctggcggg ctatattatgt 180
caagaaaggtt atggagctoa toagccttcc cttggcaaaa attacgaact tgcactggaga 240
aagttgacaa gaaagctcaa gaatacatca ggatggcagc atcattaaat gctggggaga 300
caacctacag tctggaacat gccagtgacc ttogagtggg agtgccagaaa gtgtatggagc 360
tgatagacgc ttttaagtaag aagatcttaa ccttgggctt gaaccaggac cctcccaccac 420
atccaagcaa tttgcggctg cagagaatga tcagatactc agctacactt tttgtgcag 480
aaaagtgtct tggttttgat tcaactgccaa ccaagaaca gtttgaggaa ctgaaaaaga 540
aaaggaagga ggaatggag aggaagaggg ccgtggagag acaagctgcc ctggagctcc 600
agcgaagctc tgaggaaagg cagagtgccc tggcttctcg agcgg 645

```

```

<210> 8
<211> 611
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7676140H2

```

```

<400> 8
tataaactcat atagggcgca tggggcctct agatgcatgc togaagcggc cgccaagtgtg 60
ctggaanagta aaacggagtc tcaactctgtt gccaggctgg agtcagatgg cgtgatctcgg 120
gtctactcttg acctccacct cccagtgctg gggagacaac ctacagcttg gaacatgcca 180
gtgaccttgc agtggagtg cagaaagtct atgagctgat agacgcttta agtaagaaga 240
tcttaacctc gggctgaac caggaacctc caccacaacc aagcaatttg cggctgaga 300
gaatgatcag atactcagct acactttttg tgcaggaaaa gttgcttggg ttgatgtcac 360
tgccaaccaa agaacagttc gaggaactga aaaagaaaag gaaggagaaa atggagagga 420
agagggcctg gtagagacaa gctgccctgg agtcccagc agggcttgag gaaaggcaga 480
gtggcctggc tctcagagc gccaacgggg aggtggcacc tctccgagc ggcctgcc 540
ccttgagaaa ggctgagggc tggctcccac tgtcaggag tcaggcgcag agtgaagact 600
cagaccrcgt c 611

```

```

<210> 9
<211> 556
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 6923049H1

```

```

<400> 9
ggcatctctc cgcaggggcc ctgcccctt gagaaaggct gagggtggc tcccactgtc 60
aggaggtcag gcgcagagtg aggactcaga ccgctcctc cagcagatcc acaacatcac 120
atcattcacc aggcaggcca aggcgcgggg ccgcatgat gaagtgcgca ctctgcaaga 180
gaacctgggg cagctcagg acgagtatga ccagcagag acagagaggg ccactcgact 240
gtcccggagc caggctgagg aggaggacct gcagcgggaa cagttgcaga tgttgcgtga 300
accggagttg gaacgagaaa gggagcagtt tgggtggca tccctgcaca caccgactcg 360
gtcccaggac ttcagagaaa tgcgccttt tcagctggag cccagcagag agcctgcac 420
ccacctgctc tatgctggg atctaggctc tcccagctt cccagcagca cagttccaa 480
gaaccttca gttagctcaa gtoaacccac cagagtgttg tctgggcccc agccgttggc 540
aggagcgtt acccca 556

```

```

<210> 10
<211> 915
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature

```

WO 02/059311

8/18

PCT/US01/48917

<223> Incyte ID No: 8216794J1

<400> 10

```

gatagcagtc taatagactc actatatagga aagctgttac gccctgcaggc accgggtccgg 60
aattactctc cgggtcgcagc agctcactag tcggcggcgc actgcttggc atcaaagatg 120
tatgcattga tgttatcgat ctgctgcaggc aggagctctt cctctatggc ctctcagcc 180
tctggccccc tgtcactgtc catctcaaac gggltgatac aggtgggttc ctcaaaaggg 240
ttaccaggaa ccaggggctc tgagagcctc tctggggatc gttctgtctc ctcaactgag 300
gggltaggag ctgggctgtc tgggtgaatg aatggattcc ctgccactgc tctcctctcc 360
tcgtctctt cctcgaagac attgtactct ttcaggatgc gggctgaagg gtcctaaagg 420
accctgcagc caggagagcc agtagtgccc tcttccatgg ggtggcaga ggtcttctc 480
atcaaagggg ttaagagggc gccctcctag ttgctgtggc atgctgctct ggggtaaggc 540
ctctggcca accgctgggg ccacagccac actctggtgg gttgagttga gctaaagtga 600
gggltcttgg gagctgtgct gcttgaactc ggggaagagc ctagatccaa agcataaagc 660
aaagtgggtg cgaagctccc tctcaggtct ccagctgaaa agggggcatt tctctgaagt 720
ccggaccgga gtcggtgtgt gtcaggatg ccaccgaact gtccttctcc gttccaactc 780
cgttcacgca cctctgaggc tgtctcctgg cgtctctctc cagctgctcc ggsaaggtc 840
gtggcttctt ggtctgctgc tggctttatc tctggggtg ccgggtttct tggagggggc 900
cttttccagc tgggc 915

```

<210> 11

<211> 552

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 6995895H1

<220>

<221> unsure

<222> 539, 542

<223> a, t, c, g, or other

<400> 11

```

cctgttctag tttctatctt attattcaac tgtagtggaa ataacctaaa gtgcgcaaca 60
caaaaacagca gattctctgc cctcactctg gtgcactctga gttctcacc tgccaccttc 120
cccatccatc ctgctcctgc tctgtcctgc tccactggct cctgccagac cctcggccc 180
aaaggctccc tctcactctc tggcagtcac gtgccccctc tctgcttggc cagggtgtgc 240
ttcagctccc gcagattctc tctcagcacc tctcactcat ccaggcggcc gcaactgctg 300
gcatacaaga tgtatgctct gatgttatcg atctgtctga ggaagagctc ttctctatg 360
ggctcctcag cctctggccc actgtcactg tccatctcaa aggggtgat acaggtggc 420
tctcaaaagg gttaccagg aaccagagg cttgagagcc tctgtctggc atgttctct 480
tctcactgca aggggttagc agctgagctg tctggctgaa tgaatggatt ccttgccant 540
gnttctctcc cc 552

```

<210> 12

<211> 539

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 6889439H1

<400> 12

```

ggcactgact gaccagcagt ggaagaggca cttttggccc caggggtctg gcaggagcca 60
gtggagcagg acagagggca ggcaggatgg atggggaagg tggcagggtg agaactcaga 120
tgcacacagg tgagggccag gaatctgctg ttttgggttg cgcactttga ggtatttcca 180
ctacagttga ataataaaat agaaactaga acagggagaa tcagcattca gttctgtctt 240
ttcctgttta ttattactat cttttgtaat cggaggttta ccccttttga agggacttta 300
cattttactc accagatata aactaaatgc agctctgttg gcccagggc agaatggct 360
gctgtgtacc tcttgggtcc atttctactc gcttagtctt ggttcttat gcagattat 420

```

WO 02/059311

9/18

PCT/US01/48917

agggcagcct ttttagagcc cttcctttag ccaagacaga gaagatagat tccactgagc 480
tctattctgc tctgacagaa gcccatccct agtaggctgt gaggltccatt tccactggg 539

<210> 13
<211> 409
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 70405975D1

<400> 13
ctactgtgta cctcttgggt ccatttgcta ctgcctagtc ttggttcctt atgcagtatt 60
atagggcagc ctttttagag cccttccttt agccaagaca gagaagatag attccactga 120
gctctattct gctctgacag aagttccatcc ctagttagct gtgagttcca ttccaacctg 180
ggcgcctctt cccctgctct gcacttctctg tctgtacat agaaggggga gglgctgcta 240
tgaaggggag agtttagacc caggagagcc cagcaactct ctttaggggt gggtagtgg 300
aatattttcc caggytctat ttctctagtt taagttcttt ttgtctctt ccaggaagtt 360
aagttccag tgacaggtat caatgtgaat ctggtctcta gcttttttag 409

<210> 14
<211> 698
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8117068H1

<400> 14
taagcctgca ggtaccggtc cgtgaattcc cgggtcgagc gggggagggt ctgctatgaa 60
ggagagagtt tagaccagc agagcccagc acctcttttt aagtggggt gatgggata 120
ttccaccagc ctctattttc tcagtttaag ttcttttttt tctctttcag gaagttaagc 180
tcccagtgca gggatcaat gtgaattctgg tcttgagctt tttagaaaat aagagtgtg 240
gccggggcgc ctggtccagc cctgtaatcg caacactttg ggagcccgag gcaggcagat 300
caacaagtca gtagatcgag accaccctgg ctaacacagt gagatccgct ctctgctgaa 360
aatataaaaa attagccagc cgtgtgtggt ggggtgctgt ggtcccagct acctggagg 420
ctgagccagc agaatgctgt gaaccgggga ggcggagctt gcagtgagcc gagatcgac 480
cactgtcaact ccagcctggg cagtgaagat gagactccgt ctctaaaaaa aaataagagt 540
ggttcaactt ttggctgata ttagtattat ttgcgagaag agtttccac ctctctctga 600
tagacaactt gctgtcaact ggagaaatct tctctagagc ttggccaagg ttctctgata 660
tgggtctggt taaggccgg gctctotatta ggtcggga 698

<210> 15
<211> 172
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 747646H1

<220>
<221> unsure
<222> 171
<223> a, t, c, g, or other

<400> 15
gtgglttaact ttttggttga tattagtatt attttgaga agagtttccc cacccttctt 60
ctgatagaac aattgtctgt cactggagaa atctccctcc agagotttgg caaagttact 120
ctgatctggg tctgtttaaa aggcgggtc tctaatttag gaatcgggtga nt 172

WO 02/059311

10/18

PCT/US01/48917

<210> 16
<211> 696
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7984383H1

<400> 16
tgctggtagc cctgcaggta cggtcggaaa ttcccgggtc gacgggagga atcgggtgatt 60
tggaagcttt tgcagaacat caccagaaga agggaaagctt cccagagtca gaagctaaat 120
taaaataaatt tccaaggcta ttgatcagc cttcttctt tttggttcac tgtgtcctga 180
cttggggcac tgatgagatt ttttattttt tttgagacgg agtttcact tgttgcccag 240
gctggagtggt agtggcatga tctcggctca ctgcaacctc cacctccctg gttcaagtga 300
tatttctgcc tcagcccctt gaatagctgg gattacaggg gtggggccacc acgctcgtt 360
aatctctgta tttttgtag aggtagagtt tcacctttt ggcagggctg gtttcgaact 420
cctggcctca agtgcctgc cacctggccc tcccagaagt ctgggaatca aggcgtgagc 480
cacctgtccc gccctgatga gatattttat tcccaatggt agtattgaga aactgaaatg 540
tttgaagaag caaacaccag atcgtgctgg tagcaccaca gtaactaaac tggttggtca 600
attaaggcca gaaaggaaa tcgtgtattt agctctgggt ctttggttta caggaacata 660
actcttaagt gacatctggc gtttgataag ccattt 696

<210> 17
<211> 711
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8073132J1

<400> 17
agtgtgctgg aaccgattat gtatggaaag cagcaattag cttctggaac tacctgottt 60
tcagttcact cttcttctcag agaaatgaga ccttaagcag ggaaggacct actggccatc 120
tgagaaaccc tctagagtaa cagagaaatg tcatgggctt cagggaaaga aagggcctg 180
taccaggcaa cactcttagat gcctggttgc caacttctca aaacaaccca taacaaccag 240
caggtgtggt aactcggctga catcatttaa cacactcaata tggctgtttg tgaacagatc 300
ccaaatgcaa agcgtccttc acacacttct acccactccc catctccaca cccctatada 360
agatcttttc attcaaccaa aaggataaga aatgtgtac agaaatgta acatcaaac 420
aaatacagat tatttlttca tcaagagagt taaaacattc ccattatca gaggggtcctt 480
caccaaatg taacaaatg aatttcagct ttcccaccag taaccagta aagtatggcc 540
tagagctagt taccggcat aaagtacta gtggctcagt attgttcatt cagttcagat 600
yagtaactgc agaaactcta cccagggct gagcacat ggcctatcat gatgtoggat 660
gtccgttaag agtaatgttc ctgtaacaaa agcaccagag ctaattacat t 711

<210> 18
<211> 528
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6744627H1

<400> 18
gatttctgta gatgttcaca tttctgtgac acagtttctt atccttttgg ttgaatgaaa 60
agatcttgta taggggtgtg gagatgggga gtgggtagaa gtgtgtgaa gacgctttgc 120
atttgggatc tgtacacaaa cagccatag agtgtgttaa tgaatgtcag ccagttacca 180
accctgctgg ttgttatggg ttgttttag aagtggcaa ccaggctctc aagatgttgc 240
ctggtacagg cctcttttct tcctgagggc ccatgacatt tctctgttac tctagaggg 300
ttctcagatg gccagtggc tctctcctgc ttagggtctc atttctctga aaagaggatg 360
aactgaaaag caggtagtct cagaagctaa ttgctgctt ccatcataat tattttctt 420

WO 02/059311

PCT/US01/48917

11/18

```

gtgagaacat ttctctttaa attagctagt gatttttgatt aagactaatt cactaaacat 480
aaccttccct ccaatccact caggtagcaa tctgtacgta actaaaag 528

```

```

<210> 19
<211> 683
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 6749637H1

```

```

<400> 19
tgaatgaatt aggtgeracc tatagaagag ctatgacgtc gcatgcaagc gtacgtaagg 60
ctcgggaattc ggctcgagaa agcaggtagt tccagaagct aattgctgct ttccatcata 120
atctatctttc ttgtcgagac atttctcttl taallagcta gtgattttga ttaagactaa 180
ttcactaaac atacccttc ctaaatcnc ctccagtagc aatctgaag taactaaaag 240
cattgaaaaa cacacaagaa acatttttaa aaactatttl taaggcctg gccgggtgca 300
gtggctcact cctgtaatcc cagcaacttg ggagcccgag acygglygat cacctgaggt 360
caggaatttg agactagcat ggccaacatg gtgaacccc gtgtctacta aaaatacaaa 420
aattagctgg cgtggtggc agacgcctgt aatcccagct actcagggag gctgaggcac 480
gagaatcgct tgaacctggg agcgggaggt tgcagtgagc caagattgag ccactgcact 540
ccaactcggg tgacaagagc aagactcctg ctcaaaaaa acaaaaaaa ctgactgatt 600
gatatatact aacaaaacta aaatcatatt ctttttgatta agtttatcca tggctgtatt 660
ctctatgaa ttcttccatt taa 683

```

```

<210> 20
<211> 339
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 5073165F8

```

```

<400> 20
ccaactaaa atcatattct ttgtattaag tttatccatg gctgtattct tctatgaatt 60
cttccattta atttctccac gattatcttc tctgttaaat acatcacagc agttagaatt 120
ctctacccat cagctgtacc atgtgcgaga attcctgcag gcaaaaagt ggagttacag 180
agatgggttg acagcaggca aactttggcc tatgtattat aaccacaact tcaagttctt 240
acctcatgag aatattccac ctctctttag tcttcaagg caaacagccc cgtctcata 300
ccagatgagc aaggtcttga tatggcatag cagatctcc 339

```

```

<210> 21
<211> 494
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 6441214H1

```

```

<400> 21
tctaccalc agctgtacca tctgcagaa ttcctgcagg cacaaagttg gagttacaga 60
gatgggttga cagcagcaaa acttggccta tgtattalaa caacaacttc aagttcttac 120
ctcatgtgaa tattccact ttctttagtc ttccaagcca aacagcccgc tctcatcacc 180
agatcagcaa ggtcttgata tggcatagca gatctcceta gacacagatc atgagaaaag 240
atggagagca cttagggatt caggcatcag atgaagttgg cltttccctt ttatgccttg 300
tttgtattta cctgtctaa tacaataagg atacttaact attgtacttg cagctcaata 360
tctctttgct gttcagatac taaaatgtac ctctgagcca ttgtgagctg tgtggtaggt 420
tggacattgg catagttggt gatgggactc aaaaatgaaa ggtggtctct ttaccaggtc 480
acagactglt gcag 494

```

WO 02/059311

12/18

PCT/US01/48917

<210> 22
<211> 226
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2436362H1

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

<400> 22
gancagtttt acagaagtat agtgcagatc ccttaakgat cctatcatcc cagtttcaca 60
ctggaganaa ctggagctca gagattgagg aacotttga ctgccatctg tgcagtcaac 120
caactctgca tgaccacatt tccctccatt agcactacca gcattgctgt gaggtagtta 180
ctcagctgtt cacttgacc cagacgtaga aaagttaaga gngatt 226

<210> 23
<211> 606
<212> DNA
<213> Rattus norvegicus

<220>
<221> misc_feature
<223> Incyte ID No: 702073518T2

<400> 23
atctgatcat tctctgtagc cyaagagtdt ttggatgtgg tgaagggtct tggttcaagc 60
ccaaggttaa gatctcttta cttaaagcat ctattagctc gtacaccttc tgcacttcca 120
ctcgaaggtc gttggcgtgc tccagactgt aggtctctcc occagcattt aatgatgctg 180
ccatctctgt gtattcagga gctttctgat caactttctc catgcaaatg cgaagtcttc 240
cgtagagctt cagcatgca ggcgtgtgct ccttctcatc catctgtgct tctctcttga 300
gcagcgtgct cttgcagtc gtgcagcagc ggcctggctc atcatcttc tgcctcagga 360
ctgagctgac actgctcatg ctactgatgc tgcctggcgg ggaagcattg acactgttag 420
gtgactggct gggcctgggt tgggtgctca gggagtcctt gctggcactg ytgagcttat 480
ttgccaaagg caggccaatg agctccatcc acttcttcca catgatagac ccacagagga 540
ggcactgggt acgggggttc cggatgctga acttattccc acagctctga cagaatgaac 600
atctgg 606

<210> 24
<211> 556
<212> DNA
<213> Rattus norvegicus

<220>
<221> misc_feature
<223> Incyte ID No: 702549617T1

<400> 24
gctatctgat cattctctgt agcgaagag tgtttggatg tggggaaggg tottggttca 60
agcccaaggt taagatcttc ttaactaaag catctattag ctggtacact ttctgcactt 120
ccactcgaag tgcgttggcg tgcctcagac tgtaggctgt ctcccagctc atttaatgat 180
gctgccatcc tgatghalc agagccttcc tcatcaactt tctccatgca aagctogaatg 240
ttctcgtaca gcttcaagat gtcaggcgtg tgcctctctt catccatctg ctgctctctc 300
aggactgagc tgcactgct catgctactg atgctgctc ggggggagcc atggacactg 360
ttaggtgact ggcctgggct ggtgtgggtg ctccaggagc ccttctctggc actggtgagc 420
ttatttgcga agggcaggcc aatgagctcc atacacttct tgcacatgat agaccacaag 540
agggcgcagt ggtgac 556

WO 02/059311 13/18 PCT/US01/48917

<210> 25
 <211> 452
 <212> DNA
 <213> Rattus norvegicus

<220>
 <221> misc_feature
 <223> Incyte ID No: 702448734T1

<400> 25
 tatctgatca tctctctgtag cogaagagtg tttggatgtg gtgaagggtc ttggttcaag 60
 cccaaggtla agatcttctt acttaagca tctattagct cgtacacttt ctgcacttcc 120
 actcgaaggt cgttggcgtg ctccaagact taggttgtct ccccagcatt taatgatgct 180
 gccatccctga tgrattcagg agctttctga tcaactttct coactgcaaa tcgaagtttc 240
 tcttagagct tcacgatgtc aggcgtgtgc tcttctctat coactgtctg ctctctcttg 300
 agcagctgt ccttgagtg cgtgcagcag cggatgcgct catcatcctt ctctctcagg 360
 actgagctga cactgctcat gctactgat ctgctctcgg gggagccatg gacactgta 420
 ggtgactggc tggggctggg gtgggtgctc ag 452

<210> 26
 <211> 567
 <212> DNA
 <213> Rattus norvegicus

<220>
 <221> misc_feature
 <223> Incyte ID No: 702081830H1

<400> 26
 gtgctaagca aggtccatt ccattcttag caccaaagag ggaaaaaag taaaagata 60
 cctccatgt cagctagrga agaatcagcc tcaccagcca gatcatatt accactgatt 120
 cagaanaana cagcgaagag tagcttactt taatgatgct gccatcctga tgtattcagg 180
 agctttctga tcaactttct coactgcaaa tcgaagtttc tctgtagact tcacgatgtc 240
 agcgtgtgtc tcttctctat coactgtctg ctctctcttg agcagcgtg ccttgcagtg 300
 cgtgcagcag cggatgcgct catcactctt ctctctcagg actgagctga cactgctcat 360
 gctactgatg ctgcctcggc gggagccatg gacactgta ggtgactggc tggggctggg 420
 gtgggtgctc agggagctct tgcctggcact ggtgagctta ttgccaagg gcaggccaat 480
 gagctccata cacttcttgc acatgataga cccacagagg cggcagtggg gacggcggtt 540
 ccggatgctg aacttattec cacagtc 567

<210> 27
 <211> 383
 <212> DNA
 <213> Rattus norvegicus

<220>
 <221> misc_feature
 <223> Incyte ID No: 702053775T1

<400> 27
 tsagtatctg atcatctctg tagccgaaga gtgtgatgtg gtgagggtct ggtcagccca 60
 tttagatctc tacttaagca tctattactc tacactttct gtgcttccct ccgaaggctg 120
 ttgggtgct ccaactgta ggtgctctcc ccagcattta atgatgtctg catcctgatg 180
 tattcaggag ctittctgac aacttctcc atgcaaaagc gaagttcttc gttagagctc 240
 acgatgtctg gegtgtgctc ctctcctcc atctgtctgt ctctcttgag cagcgtgtcc 300
 ttgcagtctg tgcagcagc gatgcgctca tcatctctct cgtccaggac ttgagctgac 360
 actgctcatg ctactgatgc tgc 383

<210> 28
 <211> 255
 <212> DNA
 <213> Rattus norvegicus

WO 02/059311

14/18

PCT/US01/48917

```

<220>
<221> misc_feature
<223> Incyte ID No: 700883909H1

<400> 28
gttgatccctt acatgtggga acctcaagag cttgggtgcta tgagaagcca tctttctgat 60
tttaaaaaac atcgtgctgc aaggattgac cactatgta ttgaagtcaa taaattaata 120
atcaggltggg aaaagcttac cgcatttgac agaacaata ctgagacctc aaagattaga 180
gcaatagaaa agtctgtggt gccttgggtc aatgaccagg atgttccatt ctgtccagac 240
tgtgggaata agttc 255

<210> 29
<211> 250
<212> DNA
<213> Rattus norvegicus

<220>
<221> misc_feature
<223> Incyte ID No: 700883983H1

<400> 29
gttgatccctt acatgtggga acctcaagag cttgggtgcta tgagaagcca tctttctgat 60
tttaaaaaac atcgtgctgc aaggattgac cactatgta ttgaagtcaa taaattaata 120
atcaggltggg aaaagcttac cgcatttgac agaacaata ctgagacctc aaagattaga 180
gcaatagaaa agtctgtggt gccttgggtc aatgaccagg atgttccatt ctgtccagac 240
tgtgggaata 250

<210> 30
<211> 257
<212> DNA
<213> Rattus norvegicus

<220>
<221> misc_feature
<223> Incyte ID No: 701092129H1

<400> 30
gtgggtttggc atctcataka gccaatgggg acatgaggtc tcttcaagc atacctgccc 60
ccttgcgaaa ggctgagggc tggctcccac tgtcagaagg tcaggtacag agtgaagacc 120
ccgacctctc ccttcagcag atctcataaca ttacatcytt calcaggcaag gccaaagctg 180
caggccgcac agatgaggtg cgcacacttc aagagaacct gcggcagctg caagatgagt 240
atgaccagca gcagact 257

<210> 31
<211> 473
<212> DNA
<213> Rattus norvegicus

<220>
<221> misc_feature
<223> Incyte ID No: 702137890H2

<400> 31
ctatgctgc cggggctctc tgaggaactc cgaaggccgg gagtccggag ctggagctgg 60
agagtagct gccgctgac tgtgggaagt cgcagggaga gaggaaaggt gttggcttc 120
tcggagccgc cacgacagc gcagcgtgg ctgagattcc ttctgtgaa gagggaaag 180
ctatccaag gccacaagt cacttcccag ttatccaaga gctgcttgc caagtaaca 240
ctgcccagaa cctgtccct tgggagagct caagctgact tggcatgag tccctgctt 300
cctggggttg agagaagtac aactatggca tctttggatg acgcagggga agtgagggaa 360
ggctccctgt gccctctgt cctcaaggac cttcagttct tctatcaact tcagtcacat 420
tatgaggaag aacactcagg agatcgtgat gtcaaaagggc aaattaaaaa tct 473

<210> 32

```

WO 02/059311

15/18

PCT/US01/48917

<211> 433
 <212> DNA
 <213> *Macaca fascicularis*

<220>
 <221> misc_feature
 <223> Incyte ID No: 701741076T1

<400> 32
 caaagtgaca gtcactgtca gataacaagc acaatctgct acagtctgtg acctggtaaa 60
 gagaacacct ttctatttgt ggtcccatc accaactatg acgatgtaca acctactgca 120
 ccagtcacaa tgactcagag atacatttta gtatctgaac agcaaagaca tatttgagctg 180
 ccagtacaat gagtaagtat ccttagtgta tttagacaggg ggaatacaca acaaggcata 240
 aaagggaaac gcaaacctca tctgatgctt gactccctaa gtctcttcca tcttttctca 300
 tgatctgtgt ctaggagat ctgctatgcc gtatcacgac cttgctcatt tgggtgatgag 360
 atggggctat tgccttggga agactaacgg aaggggtgat attcacatga ggtaagaact 420
 tgaagttgtg gtt 433

<210> 33
 <211> 645
 <212> DNA
 <213> *Canis familiaris*

<220>
 <221> misc_feature
 <223> Incyte ID No: 702769948H2

<220>
 <221> unsure
 <222> 562
 <223> a, t, c, g, or other

<400> 33
 tgagtcttca ctctgcccct gacttctga caacgggagc cagccctcag cttttctcag 60
 ggagagcagg tcccctcga agggacacta cctcccgtt gcccgatga gatgogaggt 120
 cactccgctt ttccctcaag cgtcgttggg attctaggac agcctgtctc tccaggatcc 180
 tcttccctcc calttccctgc ttccctttct ttttcagtcc ctcaaatgtt tctctggttg 240
 gcaatgacat taacccagc aacttttctt gcacaaaaag tgtagtctgag tatctgatca 300
 tcttctgtag ccgcaaggta tttggatgtg gtggagggtc ctggttcaag cctaaggta 360
 agatctcttt acttaaaagca tctatgagct cgtacacttt ctgcaactcc actcgaaggt 420
 caccagcag ttccagactg taagtgttt ccccgcact aaggaaaaat catcgaata 480
 tcaagttc atcaggatgc ctgggaata ttgtgaaaaa ggttctlaag ggtttagggc 540
 ctacaaaag gaggcaagc thattttctc caagtagtat atttcaact cctgtataca 600
 atgcaggact ctttttttt ctaaatataa gtgtataat cggg 645

<210> 34
 <211> 542
 <212> DNA
 <213> *Canis familiaris*

<220>
 <221> misc_feature
 <223> Incyte ID No: 702245053H1

<400> 34
 cctgaactca actctgtctt gacagaagtc tgtcccatt agcctgtggg ttctttcacc 60
 tggggcttcc tctgccctgc tgggtacatc acacctgtgc agcagagggg ggaggggggg 120
 ttattgccct gaaggcaata gtttaggccc aggagagcct agcagctctt ttaagcagc 180
 ggaggggggg gtgatgtttt accagggcat ttccctattg gtttaagtcc tttttgtct 240
 ctctcaggaa gttaaacctc cagtgacagg gtatcactgg gtctgacac agcctgttta 300
 gaaaatagga gtggtagact ttttggttga tggaggcagt agtatttggg agaagaatct 360
 ctacccaatg gttcctctga atgatggaac aatttctccc cagagaaacc tccctctaga 420
 attttgtgga aaataccctg atctgggctt tttcagaggg tccaggctat aatttaggat 480

WO 02/059311

16/18

PCT/US01/48917

tggtaatcag gaagcttttg tggaaactcg ccatgggagt ggaagcttcc ttgagtcaga 540
ag 542

<210> 35
<211> 281
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: g921283

<400> 35
gaactcaag ttcttaacct atgtgaatat tcaccttttc tttagtcttc caaggcaaac 60
agccccgtct cabcaccaga tgagcaaggt cttgacatgg attagcagat ctccctagac 120
acagatcatg agaaaagatg gaagagactt agggattcag gcatcagatg aagtgtgctt 180
ttccctttta tcctttgttt gtattttacc tgcctaatca actaaggata cttactcaat 240
gtacttgtag tcaatatgt ctttgcgttt cagatactaa a 281

<210> 36
<211> 6719
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: GNFL.g8572864_000002_002.edit

<400> 36
aaattacgac ttgtcatgga gaaagttgac cagaaagctc cagaatacat caggatggca 60
gcatacattaa atgctgggga gacaacctac agtctggaac atgccagtga ccttcgagtg 120
gaagtgcaga aagtgtatga gctgatagac gctttaagta agaagatctt aaccttgggc 180
ttgaaccagg acctccacc acatccaagc aatttgcggc tgcagagaat gatcagatac 240
tcagctacac tttttgtgca ggaagaattg cttggtttga tgcactgcc aaccaaaaga 300
cagtttgagg aactgaaaaa gaaaaggaag gaggaaatgg agaggaagag ggcctggag 360
agacaagctg cctggagtc ccagcgaagg cttgaggaaa ggcagagtg cctggcttct 420
cgagcgcca acggggaggt ggcactcttc cgcaggggccc ctgcccctt gagaaaggct 480
gagggctggc tcccactgtc aggaggtcag gggcagagtg aggactcaga cccgctcttc 540
cagcagatcc acaacatcac atcatlcatc aggcagggcca aggtcgggg ccgcctggat 600
gaagtggaga ctctgcaga gaaactgcgg cagctgcagg accgatatga ccagcagcag 660
acagagaagg ccactcgagct gtcccggagg cagctgcagg agggagacct gcagcgggaa 720
cagctgcaga tgttgctgga acgggagttg gaacgagaaa gggagcagtt tcgggtggca 780
tcctcccaca caccgactgc gtcctggac ttcagagaaa tcggcccttt tcagctggag 840
cccagcagag agcctcgcac ccaacctgtc tatgctttgg atctaggctc ttcccagtt 900
ccaagcaga cagctcccaa gaccccttca cttagctcaa ctcaaccac cagagtgtg 960
tctgggccc cagccgttgg ccaggagcgc ttaccocaga gcagcatgcc acagacaat 1020
gaggggccct ccttaaaccc ctttgatgag gaagacctct ccagcccact ggaagagccc 1080
actactggtc ctctgctgc aggggtttcc ttgaccctt cagcccgat cctgaaagag 1140
tacaatcctt tcgaggaaga ggcagggagg gaggagcag tggcagggaa tccattcaat 1200
cagccagaca gccagctcc taacccttc agtgaggag acaaacatcc ccagcagagg 1260
ctctcagacc ctctgggtcc tggtaacccc ttgagggaac ccaoctygt caacccttt 1320
gagatggaca gtgacagtg gccagagct gaggagccc tagaggaaga gctcctctg 1380
cagcagatcg ataacafoaa ggcatacacc tttagtgcca agcagtgcg ccgctggat 1440
gaggtagagtg tgcagcaga gaatctgcgg gactggaac acaccctgccc caagcagaag 1500
gggggcactg actgaccagc agtgagagag gcaaccttgg gccagggyl ctggcaggag 1560
ccagtggagc acgacagag gcaaggcaga tggatgggga agtgccagc gtgagaactc 1620
agatcacac agtgaggggg caggaatctg ctgttttgtg ttgcccactl tgaggtatt 1680
ccactacagt tgaataataa aatgaaact agaacagga gaatcagcat tcagttgctg 1740
cttttccctg ttattattac tatctttgt aatcggaggt ttacccttt tgaagggact 1800
ttacatlttt actaccagga tataactaaa tgcagctctg ttggccocag ggcagaaatg 1860
gtcgtgtgt actccttggg tccatttgt actgcctagt cttggttct tatgcagtat 1920
tatagggcag cctttttaga gcccttccct tagccaagac agagaagata gattccactg 1980
agctctattc tgcctgaca gaagtccatc cctagtaggc tgtgagttcc atttccactg 2040

WO 02/059311

PCT/US01/48917

17/18

gggcgcgcctc tccocctgctc tgcacttctc gtctgtacaa tagaaggggg aggtgctgct 2100
atgagggaga gattttagac caaggagac ccagcacctc ttttlaagg ggggtagatg 2150
gaataattca caagggctca ttttctcagt ttaagttcct ttttcaagg ttcaggagc 2200
taagctccca gtgcaggta tcaatgtgaa tctggtcctg agctttttag aaaaatagag 2280
tggtggcccg gcgcgctggc tcaagcctgt aatcccaaca ctttgggag ccgagccag 2340
cagatcaaca ggtoacagga togagaccac cctggctaac acagtgaaat cccgctcta 2400
ctaaaaatad aaaaaattag caagcctggg tggtaggtgc ctgtagtccc agctactttg 2460
gaggctgagg caggagaagt ctgtgaaccc gggaggcaga gcttgcagtg agccagatc 2520
gcaccaatgc actccagcct gggcagtgag agtgagactc cgtctcaaaa aaaaataaga 2580
gtggtttaact ttttggttga tattagattt atttgtgaga agagtttccc cccctctct 2640
ctgatagaac aattgtctgt cactggagaa atctccctcc agagcttttg caaagttact 2700
ctgatclggg tctgtttaaa agccgggctc tctaatttag gaatcgttga tttggaagct 2760
tttgcagaac atccaccaga gaagggaaag tttccagagt cagaagctaa attaaaaaa 2820
tttccaaagg tatttgatca gctctctctc ttttgggttc attgtgtcct gacttggggc 2880
actgatgaga ttttttctt tttttgagac agagtttcaac tttgttgccc aggtcggagt 2940
gtagtgcaat gattctcagc cactgcaacc tcccctccc tgggtcaagf galacttctg 3000
cctcagcccc ctgatatgct aggatctacag gcttgggcca ccaagcctgg ttaactctg 3060
tatttttagt agaggtagag tttccactt ttggccaggc tggtttcaaa cctcctgctc 3120
caagtgatct gccaccctgg cctcccaaaq tctcgggaat acagcctgta gccactgtgc 3180
ccgccctgat gagatatttt atlaccaatg ttagtattga gaaactgaaa tggttgaga 3240
agcaacaacc aggtactgctg tggtagcccc acagacttta aactgttggf caatlaaggc 3300
cagaagggga aattgttaat ttagctctga gagctttaag ttacaggaa cataactctt 3360
aactgacatc tgacatcatg atagccatat gtgctcagct ctgcggtaga gtttctgag 3420
ttactcaatc gaactaatga acaabaactg accactagtc actttatgoc gttgtaactg 3480
ctctaggcca taacttcaact ggttactggt gtgaaagctg aaattcattt tgttaactc 3540
tggtagaagc cccctctgat aatgggaatg ttttaactct ctgatgaaa aataaatctg 3600
tatttgggtt gatgttcaaa tttctgtagc acatcttcta tctctttggt tgaatgaaa 3660
gatcttctat aggggtctg agtggggag tgggtgaaag tttgtgaaag accgtctgca 3720
tttgggatct gttcaacaac agccatagga gtgtgttaat gaatgtcagc cagttacaaa 3780
ccctctgctg ttttatgggt ttttllgaga agttggcaac caggeactca agatgttgc 3840
tggtagcagg ctcttttctt cctgaggccc catgacatll cctctgtae ctgaggggtt 3900
tctcagatgg ccagtaggct cctccctgct tagggttca tttctctgaa aagaagatga 3960
actgaaaagg aggtagttcc agaagctaat tgcgttttc catcaataat attttcttg 4020
tgagaacatt tctcttttaa ttagctagtg attttgatla agactaatc actaaaaata 4080
cccttccctc aatcaaccct aggtagcaat ctgtaagtaa ctaaaagcat tgaaaaacac 4140
acaagaacaa tttttaaaaa ctatttttaa agccctggcc ggggtcaagt gctcactcct 4200
gtaatcccag cactttggga ggcagagacg ggtggatcac ctgaggtcag gaatttgaga 4260
ctagatggcg caacatgggt aaacccctgt tctactaaaa atacaaaaat tagctggggc 4320
tggtagcaga gcctgtaat cccagctact cagggaggct gaggcacgag aatcgttga 4380
aacttggagc cggaggttgc agtgagccaa gattgggcca ctgcaactca accctgggtga 4440
caagagcagc actcctctc acaaaaaaca acaaaaactg actgallgaa tacaactaac 4500
caactcaaaa tcaattctt ttgattaagt ttatccatgg ctgtattctt ctatgaatc 4560
ttcattttaa tttctccagc attatctct cctgtaaata catcaagga gttagaatc 4620
tctaccactc agctgtacca tgtcgagaa ttcctgagc caaaaagtg aggttcaaga 4680
gatgggttga cagcaggcaaa acttggccta tgtattataa ccaacaactc aagtcttac 4740
ctcatgtgaa tattcaacct tcttttagtc ttccaaggca aacagcccng tctcatcacc 4800
agatgagcaa ggtcttgata tggcatagca gatctcccta gacacagatc atgagaaaag 4860
atggaagaga cttagggatc caggcatcag atgaagttgg cttttccctt ttatgccttg 4920
ttgtatttca cctgtctaaa taactaagg atacttactc attgtacttg cagctcaata 4980
tgtcttctgt gttcagatac taaaatgtac ctctgagctg ttgtgagctg tttgttaggt 5040
tggcaatggc catagttggt gatgggactc aaaaatgaaa ggtggtctct ttaccaggct 5100
acagactgta ccagattggt cttgttatct gacaaatgact gtcacttga ggtctgltga 5160
tttgaatgca ctactctggg gccttctgatt gtagctctct ttaaaaaaaa taaaactgta 5220
gatagaggtt gggggtctgt gtgctctggt gcaaatgagt tacaagtgag aalcaataga 5280
tgacatccct tctctttct tgatgacaac catctgagta tcaagaatag tccagccctc 5340
gtttgtttg tctggttaag cctctggaaa aaaaatgagg tcaactgggt ctgacaggg 5400
gatgatacag ggttccagtt ctgccttat tcccagttat tctgtcagtg ctggttaaat 5460
gaacagttt acagaagat agtgacatc ccttaatgat cctatcacc cagtttcaaa 5520
ctggaagaaa ctgaggtca gagattgagg aacctgtgta ctgcatctg tggcagtoac 5580
caoctctgca tgaccaact tccctccatt agcactacca gcatgcctgt gaggtagtta 5640
ctcagctgtt cactgaccc cagacgtaga aaagtttaaga gagattagcc tggcttaggc 5700
cacacagca gtcagtggtg gaatgaggtt taaaatccag atctgcctga cttaatttgc 5760
cctcagtaat aagttagctg gtcaggtagt ggtgtggtg tgggtgggtg tcaaacattg 5820

WO 02/059311

18/18

PCT/US01/48917

```

tgtcaagctc  aaagagggca  gaaaagggca  atggggaatg  aggttgsgtg  ctgctgagct  5880
ggtttcogat  agtgactgcc  tgagtctgtg  aaactgagac  ctcttaggg  ccttaagat  5940
atatagaltc  ttaaacccag  agagtataat  atctagtgtc  cactgtagge  tagactgagt  6000
gctaagtatc  accgatggtc  tacctcagtc  ttcccacaa  cccatttaca  catgcagaaa  6060
tgagtgagtg  gtgaagttaa  gettgcttag  ggttgcactg  ctgggttcc  agoctaggtc  6120
tgtctgactg  caaaaccocaa  getgccttag  attgtgttgt  atgtggcac  ccatgtttct  6180
aatgtggaat  ggggtgttcc  atccctagca  agtgcctct  gtggactttt  aagatgaaaa  6240
acttctgctc  aactcttcag  gaagcctagc  gtagtgttaa  aaatacttca  gaaagcactt  6300
ctggttttag  ctacagacag  tagagtttga  aagttatcct  atccttacca  caagaaaaag  6360
ctgggcaaac  tcaaaagcat  tagcctttct  tggctccatt  agagagctga  gtgtccaggt  6420
caaaactcca  tccaaaaatc  tagagagagg  tgagtccaga  gagtgagcaa  gatctaattg  6480
cctggagcag  aagccacggg  agccatagct  agtaggaaaa  cttaaalgtt  aattttgatg  6540
aactgctgga  ggttgagtdt  ggaactagcat  gggagggagc  aactccocct  caccoccaac  6600
gcttccctgg  gtttaacct  acttggttct  attggtagag  agctgaggaa  gacctctctg  6660
tgactctggc  aggggaaagg  gagaatcacc  tttgtaatca  gagccttctc  cataataaa  6719

```

【国際公開パンフレット(コレクトバージョン)】

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

(19) World Intellectual Property Organization
International Bureau



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

PCT

(10) International Publication Number
WO 02/059311 A3

- (51) International Patent Classification⁷: C12N 15/12, BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US).
- (21) International Application Number: PCT/US01/48917 (74) Agents: STREETER, David, G. et al.; Incey Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (22) International Filing Date: 13 December 2001 (13.12.2001) (81) Designated States (national): AI, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GH, GM, GR, HU, 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: 09/740,235 18 December 2000 (18.12.2000) US
- (71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (72) Inventors: and
- (75) Inventors/Applicants (for US only): LASEK, Amy, K., W. [US/US]; 237 41st Street, #5, Oakland, CA 94611 (US); KRASNOW, Randi [US/US]; 817 Santa Fe Avenue, Stanford, CA 94305 (US); WARREN, Bridget, A. [US/US]; 10130 Parkwood Drive #2, Cupertino, CA 95014 (US).
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BI, CH, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
with international search report

[Continued on next page]

(54) Title: COLON CANCER MARKER

```

          9       18       27       36       45       54
5'  GCC TGC CGA GTT CCG AGC GAC CGA TGG AGA TGG CGG CTG CCG CTG AGT GAC GGA
          63       72       81       90       99      108
CGG TGG AGG CCC AGA GCC CCG GCC TGA AGG GGG GGA CAA ACC TGG GTG CCC GCA
          117      126      135      144      153      162
GGA GCC CCG CAG GGT GTC TTA CAA GTA TCA AGA ACT TAC TAT ATG TGG TTG AAT
          171      180      189      198      207      216
AAA CAA TCA AGG TAA AGA GCA TCA AAT AAA AAC TTC TGC TTG ATG ATA AGT ACT
          225      234      243      252      261      270
TCA GAC AAT CCC CCA GTG GCT GAA GTG GCA TAT GAA TTA TGA AGT TGG ATC AAT
          279      288      297      306      315      324
TGG AAT GAA TGT AAG AGA AAT GCC AAG GGC TCC TCC TAC TCC AGA GAG GAA ACC
          333      342      351      360      369      378
TCA TCC AGC GCC ATG AAG CCA CTT CCT CAC CAT CTG TGT GCT GCT TAA GCT AAT
          387      396      405      414      423      432
GCT CCG GGA ACC ATG GTT CCT TGG GAG GAA TCA AGC TGA CTC TTG GCA TGA GAT

```

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

WO 02/059311 A3

WO 02/059311 A3 

(88) Date of publication of the international search report: 10 April 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 01/48917
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/12 C07K14/47 C12Q1/68		
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		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) BIOSIS, EPO-Internal, WPI Data, CHEM ABS Data, EMBASE, EMBL		
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! EMBL; 28 November 2000 (2000-11-28) NIELSEN E. ET AL.: "Homo sapiens FYVE-finger containing Rab5 effector protein Rabenosyn-5 mRNA" retrieved from EMBL, accession no. AY009133 Database accession no. AY009133 XP002225162 see enclosed protein sequence which shows 99.7% identity with SEQ ID N°1 in 784 aa overlap --- -/-	1-3,13
<input checked="" 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:		
<p>*A* document defining the general state of the art which is not considered to be of particular relevance</p> <p>*E* earlier document but published on or after the international filing date</p> <p>*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)</p> <p>*O* document referring to an oral disclosure, use, exhibition or other means</p> <p>*P* document published prior to the international filing date but later than the priority date claimed</p> <p>**I* 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</p> <p>**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</p> <p>**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.</p> <p>*G* document member of the same patent family</p>		
Date of the actual completion of the international search	Date of mailing of the international search report	
16 December 2002	10/01/2003	
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentstrasse 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Vix, O	

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

INTERNATIONAL SEARCH REPORT

 Int. Application No.
 PCT/US 01/48917

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	NIELSEN ERIK ET AL: "Rab5 effector, is complexed with hVPS45 and recruited to endosomes through a FYVE finger domain." JOURNAL OF CELL BIOLOGY, 'Online! vol. 151, no. 3, 30 October 2000 (2000-10-30), pages 601-612, XP002225144 ISSN: 0021-9525 see Material and Methods: antibodies, plasmids, Rabenosyn-5 cloning and cell transfection,... in combination with the EMBL sequence entry AY009133 -----	1-20
A	NIMMRICH INKO ET AL: "Seven genes that are differentially transcribed in colorectal tumor cell lines." CANCER LETTERS, vol. 160, no. 1, 10 November 2000 (2000-11-10), pages 37-43, XP002225145 ISSN: 0304-3835 the whole document -----	1-29
A	WO 96 39541 A (HUMAN GENOME SCIENCES INC ;LI YI (US); SOPPET DANIEL R (US); DILLO) 12 December 1996 (1996-12-12) the whole document -----	1-29
A	WO 94 25625 A (CHAPELLE ALBERT DE ;UNIV JOHNS HOPKINS (US)) 10 November 1994 (1994-11-10) the whole document -----	1-29

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

INTERNATIONAL SEARCH REPORT
Information on patent family members

In International Application No
PCT/US 01/48917

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9639541	A	12-12-1996	CA 2221795 A1 12-12-1996
			CN 1194009 A 23-09-1998
			WO 9639541 A1 12-12-1996
			AU 711346 B2 14-10-1999
			AU 2818095 A 24-12-1996
			EP 0833948 A1 08-04-1998

WO 9425625	A	10-11-1994	US 5492808 A 20-02-1996
			EP 0698123 A1 28-02-1996
			JP 8509616 T 15-10-1996
			WO 9425625 A1 10-11-1994
			US 5693470 A 02-12-1997
			US 5591826 A 07-01-1997
			US 5837443 A 17-11-1998
			US 5871925 A 16-02-1999

フロントページの続き

(51) Int.Cl. ⁷	F I	テーマコード(参考)
C 1 2 N 1/19	C 1 2 N 1/19	4 H 0 4 5
C 1 2 N 1/21	C 1 2 N 1/21	
C 1 2 N 5/10	C 1 2 P 21/02	C
C 1 2 N 15/02	C 1 2 P 21/08	
C 1 2 P 21/02	C 1 2 Q 1/68	A
C 1 2 P 21/08	G 0 1 N 33/15	Z
C 1 2 Q 1/68	G 0 1 N 33/50	Z
G 0 1 N 33/15	G 0 1 N 33/53	D
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	G 0 1 N 33/574	A
G 0 1 N 33/574	G 0 1 N 33/58	Z
G 0 1 N 33/58	C 1 2 N 5/00	A
	C 1 2 N 15/00	C

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

(72) 発明者 クラスノー、ランディ

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

(72) 発明者 ワレン、ブリジット、エイ

アメリカ合衆国カリフォルニア州 9 5 0 1 4 ・ クーペルティノー ・ # 2 ・ パークウッドドライブ
1 0 1 3 0

(72) 発明者 ボーゲン、マライア・アール

アメリカ合衆国カリフォルニア州 9 4 5 7 7 ・ サンレアンドロ ・ サンティアゴロード 1 4 2 4 4

F ターム(参考) 2G045 AA40 BB14 BB50 DA13 FA11 FB02 FB07 FB08 FB12
4B024 AA01 AA12 BA36 CA04 CA05 CA11 DA02 DA06 HA14
4B063 QA13 QA19 QQ43 QQ52 QR32 QR35 QR55 QS34 QX01
4B064 AG27 AG31 CA10 CA19 CA20 CC24 CE12 DA05 DA14
4B065 AA26X AA90X AA93Y AB01 CA24 CA44 CA46
4H045 AA10 AA11 AA20 AA30 BA10 CA41 DA76 DA86 EA28 EA51
FA74 GA26

专利名称(译)	结肠癌标记物		
公开(公告)号	JP2004537971A	公开(公告)日	2004-12-24
申请号	JP2002559796	申请日	2001-12-13
[标]申请(专利权)人(译)	洞察Genomics公司		
申请(专利权)人(译)	洞察基因组公司		
[标]发明人	ラセックエイミーケイダブリュ クラスノーランディ ワレンブリジットエイ ボーグンマライアール		
发明人	ラセック、エイミー・ケイ・ダブリュ クラスノー、ランディ ワレン、ブリジット、エイ ボーグン、マライア・アール		
IPC分类号	G01N33/50 A61K38/00 C07K1/02 C07K14/47 C07K14/82 C07K16/32 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N15/02 C12N15/09 C12N15/12 C12P21/02 C12P21/08 C12Q1/68 G01N33/15 G01N33 /53 G01N33/566 G01N33/574 G01N33/58		
CPC分类号	C07K14/4748 A01K2217/05 A01K2217/075 A61K38/00 C12Q1/6886 C12Q2600/158		
FI分类号	C12N15/00.ZNA.A C07K1/02 C07K14/82 C07K16/32 C12N1/15 C12N1/19 C12N1/21 C12P21/02.C C12P21/08 C12Q1/68.A G01N33/15.Z G01N33/50.Z G01N33/53.D G01N33/53.M G01N33/566 G01N33 /574.A G01N33/58.Z C12N5/00.A C12N15/00.C		
F-TERM分类号	2G045/AA40 2G045/BB14 2G045/BB50 2G045/DA13 2G045/FA11 2G045/FB02 2G045/FB07 2G045 /FB08 2G045/FB12 4B024/AA01 4B024/AA12 4B024/BA36 4B024/CA04 4B024/CA05 4B024/CA11 4B024/DA02 4B024/DA06 4B024/HA14 4B063/QA13 4B063/QA19 4B063/QQ43 4B063/QQ52 4B063 /QR32 4B063/QR35 4B063/QR55 4B063/QS34 4B063/QX01 4B064/AG27 4B064/AG31 4B064/CA10 4B064/CA19 4B064/CA20 4B064/CC24 4B064/CE12 4B064/DA05 4B064/DA14 4B065/AA26X 4B065 /AA90X 4B065/AA93Y 4B065/AB01 4B065/CA24 4B065/CA44 4B065/CA46 4H045/AA10 4H045/AA11 4H045/AA20 4H045/AA30 4H045/BA10 4H045/CA41 4H045/DA76 4H045/DA86 4H045/EA28 4H045 /EA51 4H045/FA74 4H045/GA26		
优先权	09/740235 2000-12-18 US		
外部链接	Espacenet		

摘要(译)

本发明提供了编码结肠直肠癌标志物的cDNA。本发明还提供了用于cDNA，其互补的片段的利用率，并提供了一种突变体，用于利用所编码的蛋白质，特别是诊断和治疗结肠疾病，如结肠癌和息肉对于自己部分和抗体。本发明还提供了用于产生蛋白质和遗传重组模型系统的表达载体和宿主细胞。

number

DE CYS CQ C6

(SUS)

-102	400	Hamm,曹恩涛,West,DeS783	Hamm,曹恩涛,Conce,De3811
-107	544	Hamm,曹恩涛,West,DeS783	Hamm,曹恩涛,Conce,De3877
-108	0	Hamm,曹恩涛,West,DeS783	Hamm,曹恩涛,Conce,De3878
-101	1846	Hamm,曹恩涛,West,DeS783	Hamm,曹恩涛,Polg,De3876
-141	518	Hamm,曹恩涛,West,DeS783	Hamm,曹恩涛,Polg,De3874
-229	60	Hamm,曹恩涛,West,AmHercA,DeS783	Hamm,曹恩涛,AmHercA,De3849
-104	1176	Hamm,曹恩涛,West,Conce,DeS897	Hamm,曹恩涛,Conce,De3847
-004	141	Hamm,曹恩涛,West,AmHercA,DeS783	Hamm,曹恩涛,AmHercA,DeS880
018	528	Hamm,曹恩涛,AmHercA,DeS783	Hamm,曹恩涛,Polg,AmHercA,De3883
