

(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2005-505232

(P2005-505232A)

(43) 公表日 平成17年2月24日(2005.2.24)

(51) Int. Cl. <sup>7</sup>	F I	テーマコード (参考)
C 1 2 N 15/09	C 1 2 N 15/00 Z N A A	2 G O 4 5
A 6 1 K 39/395	A 6 1 K 39/395 E	4 B O 2 4
A 6 1 P 35/00	A 6 1 K 39/395 T	4 B O 6 3
C O 7 K 1/22	A 6 1 P 35/00	4 B O 6 4
C O 7 K 14/47	C O 7 K 1/22	4 B O 6 5
	審査請求 未請求 予備審査請求 有 (全 139 頁)	最終頁に続く

(21) 出願番号 特願2002-548128 (P2002-548128)  
 (86) (22) 出願日 平成13年12月3日 (2001.12.3)  
 (85) 翻訳文提出日 平成15年6月2日 (2003.6.2)  
 (86) 国際出願番号 PCT/US2001/046767  
 (87) 国際公開番号 W02002/046411  
 (87) 国際公開日 平成14年6月13日 (2002.6.13)  
 (31) 優先権主張番号 09/729,454  
 (32) 優先日 平成12年12月4日 (2000.12.4)  
 (33) 優先権主張国 米国 (US)

(71) 出願人 301005050  
 インサイト・ゲノミックス・インコーポレ  
 イテッド  
 アメリカ合衆国カリフォルニア州9430  
 4・パロアルト・ポータードライブ 31  
 60  
 (74) 代理人 100089266  
 弁理士 大島 陽一  
 (72) 発明者 ラセク、アミー・ダブリュ  
 アメリカ合衆国カリフォルニア州9461  
 1・オークランド・#5・フォーティファ  
 ーストストリート 237

最終頁に続く

(54) 【発明の名称】 腸タンパク質

(57) 【要約】

本発明は哺乳類の腸タンパク質をコードするcDNAを提供する。本発明はまた、cDNAの利用のために、自身の断片、補体、及び変異体を提供し、コードされたタンパク質の利用のために、とりわけ結腸癌、クーロン病、及び潰瘍性大腸炎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'  A TCC AAG CTA AAG GTA AAC AAA CTG CAA CTT ATA TCT GCA APT TAT TTT GGT
          63       72       81       90       99       108
ATA GAC AAG AGG TAT GCC AGT AGC ACA CTG GTG GCT TCA GAA GAA ATT CTC AAC
          117      126      135      144      153      162
ACC TAG CTC GCC AGA GAG TCT ATG TAT GGG ATT GAA CAA TCT GTA AAC TAA AGG
          171      180      189      198      207      216
ATC CTA ATC ATG AAA ATA AGT ATG ATA AAT TAT AAG TCA CTA TTG GCA CTG TTG
          M K I S H I N Y K S L L A L L
          225      234      243      252      261      270
TTT ATA TTA GCC TCC TGG ATC ATT TTT ACA GTT TTC CAG AAC TCC ACA AAG GTT
          F I L A S W I I F T V F Q N S T K V
          279      288      297      306      315      324
TGG TCF GCT CTA AAC TTA TCC ATC TCC CTC CAT TAC TGG AAC AAC TCC ACA AAG
          W S A L N L S I S L H Y W N N S T K
    
```

## 【特許請求の範囲】

## 【請求項 1】

哺乳動物タンパク質若しくはその一部をコードする単離された哺乳動物cDNA若しくはその断片であって、

(a) SEQ ID NO:1及びSEQ ID NO:2(配列番号1及び2)のアミノ酸配列と

(b) SEQ ID NO:1若しくはSEQ ID NO:2のアミノ酸配列に対して少なくとも80%の同一性を有する変異体と、

(c) SEQ ID NO:1若しくはSEQ ID NO:2の抗原性エピトープと、

(d) SEQ ID NO:1若しくはSEQ ID NO:2のオリゴペプチドと、

(e) SEQ ID NO:1若しくはSEQ ID NO:2の生物学的活性部位とから選択される単離された哺乳動物cDNA若しくはその断片。 10

## 【請求項 2】

SEQ ID NO:1若しくはSEQ ID NO:2の哺乳動物タンパク質をコードする単離された哺乳動物cDNA。

## 【請求項 3】

単離された哺乳動物cDNA若しくはその補体であって、

(a) SEQ ID NO:3及びSEQ ID NO:10の核酸配列と、

(b) SEQ ID NO:4乃至9及びSEQ ID NO:11乃至15から選択された断片と、

(c) SEQ ID NO:3若しくはSEQ ID NO:10の核酸配列に対して少なくとも80%の同一性を有するSEQ ID NO:16乃至29から選択された変異体と、 20

(d) SEQ ID NO:3乃至29のオリゴヌクレオチドとから選択される単離された哺乳動物cDNA若しくはその補体。

## 【請求項 4】

前記請求項1のcDNA若しくはその補体を含む組成物。

## 【請求項 5】

前記請求項1のcDNA若しくはその補体を含む基質。

## 【請求項 6】

前記請求項1のcDNA若しくはその補体を含むプローブ。

## 【請求項 7】

前記請求項1のcDNAを含むベクター。 30

## 【請求項 8】

前記請求項7のベクターを含む宿主細胞。

## 【請求項 9】

タンパク質生成方法であって、

(a) タンパク質の発現条件下で、前記請求項8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 10】

前記請求項7のベクターを含む遺伝子組換え細胞株若しくは生物。

## 【請求項 11】

サンプル中の核酸の前記特異発現を検出するのにcDNAを用いる方法であって、 40

(a) 前記核酸に前記請求項6のプローブ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の前記特異発現を示すことを特徴とする過程とを含む方法。

## 【請求項 12】

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11に記載の方法。

## 【請求項 13】

cDNAの特異発現の検出が、とりわけ結腸癌、クーロン病、及び潰瘍性大腸炎のような結腸 50

疾患の診断であることを特徴とする請求項 1 1 に記載の方法。

【請求項 1 4】

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 1 5】

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 1 4 に記載の方法。 10

【請求項 1 6】

精製された哺乳動物タンパク質若しくはその一部であって、  
( a ) SEQ ID NO:1 及び SEQ ID NO:2 ( 配列番号 1 及び 2 ) のアミノ酸配列と、  
( b ) SEQ ID NO:1 若しくは SEQ ID NO:2 のアミノ酸配列に対して少なくとも 80% の同一性を有する変異体と、  
( c ) SEQ ID NO:1 若しくは SEQ ID NO:2 の抗原性エピトープと、  
( d ) SEQ ID NO:1 若しくは SEQ ID NO:2 のオリゴペプチドと、  
( e ) SEQ ID NO:1 若しくは SEQ ID NO:2 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。 20

【請求項 1 7】

前記請求項 1 6 のタンパク質を含む組成物。

【請求項 1 8】

少なくとも一つのリガンドを同定するべく、複数の分子若しくは混合物をスクリーニングするためにタンパク質を用いる方法であって、特異的な結合を可能とするための条件下で、前記請求項 1 6 のタンパク質と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 9】 30

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 8 に記載の方法。

【請求項 2 0】

SEQ ID NO:1 若しくは SEQ ID NO:2 のアミノ酸配列を含む精製された哺乳動物タンパク質。

【請求項 2 1】

前記請求項 2 0 のタンパク質に特異的に結合する精製された抗体。

【請求項 2 2】

前記請求項 2 1 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、 40

( a ) 抗体反応を誘発する条件下で SEQ ID NO:1 若しくは SEQ ID NO:2 のタンパク質を有する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な結合が可能となるような条件下で、前記基質と単離した抗体とを接触させる過程と、  
( e ) 前記抗体を前記タンパク質より解離させ、それによって精製された抗体を獲得する過程とを含む方法。

【請求項 2 3】

前記請求項 2 2 の方法によって生成された多クローン性抗体。 50

## 【請求項 24】

前期請求項 21 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 若しくは SEQ ID NO:2 のタンパク質を有する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。

10

## 【請求項 25】

前記請求項 24 の方法によって生成された単クローン性抗体。

## 【請求項 26】

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 21 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。

20

## 【請求項 27】

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 21 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混合体形成を検出する過程とを有する方法。

## 【請求項 28】

複合体形成が基準物質と比較され、また、特に結腸癌、クローン病、及び潰瘍性大腸炎のような結腸疾患の診断であることを特徴とする請求項 27 の方法。

## 【請求項 29】

請求項 21 の抗体及び標識部分を含む組成物。

30

## 【請求項 30】

請求項 21 の抗体及び薬剤を含む組成物。

## 【請求項 31】

結腸疾患の治療方法であって、

前記請求項 31 に記載の成分をそのような治療を必要とする患者に投与する過程を含む結腸疾患の治療方法。

## 【発明の詳細な説明】

## 【技術分野】

## 【0001】

本発明は、ほ乳類の腸タンパク質をコードする cDNA に関し、また、とりわけ結腸癌、クローン病、及び潰瘍性大腸炎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。

40

## 【背景技術】

## 【0002】

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

50

にする。

【0003】

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)。結腸癌は男女でおよそ等しい頻度で発生する唯一の癌である。また米国では結腸癌の診断5年後の生存率は約55%である(Riesら (1990) National Institutes of Health, DHHS Publ No. (NIH)90-2789)。

【0004】

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

【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。とりわけ結腸癌では、これら変化が、結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、組織学的に結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)。

【0006】

家族性腺腫様ポリープ(FAP)は、結腸癌に先立ち、腺腫様ポリープ大腸菌(APC)遺伝子中で遺伝子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腺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、アディージョン、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変化は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遺伝子の遺伝は微量なイベントであるが、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、一般人の大部分の結腸癌にとって中心的であると信じられている。

【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

10

20

30

40

50

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。

【0008】

潰瘍性大腸炎が、結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。

【0009】

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。

10

【0010】

ほとんどすべての結腸癌は、エストロゲン受容体(ER)遺伝子が不活性の細胞から発生する。ER遺伝子転写の不活性化は年齢に関係しており、ER遺伝子の過剰メチル化に関連する(Issaら(1994) Nature Genetics 7:A536-540)。培養した結腸癌細胞に外来性ER遺伝子を導入することにより、著しく成長が抑制される。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。

【0011】

通常発生の間、腸上皮の形態学的変化と生化学的分化は、多数の遺伝子の表現における変化を伴う。腸上皮はリーベルキューン(Lieberkuhn)及び絨毛(villi)のクリプト(crypt)へと細分化される。哺乳動物では、成人へと変わる間、クリプトが深くなり、クリプト細胞に対する絨毛の比率が減少し、腸細胞のターンオーバー率が増加する。加えて、成人の食餌への移行は、様々な酵素レベルにおける変化によって達成される。酵素には、ラクターゼフロリジン水解酵素、リゾリーマル酸ヒドロラーゼ、スクラーゼ イソマルターゼ、アルカリホスファターゼ及びアミノペプチダーゼが含まれる。Boll(1993; J Biol Chem 268:12901-12911)らは、減法ハイブリダイゼーション法を用い、成人の腸内では発現するが幼児の腸内では発現しないウサギの遺伝子のためにcDNAクローンを分離した。彼らは腸の発育とセル分化のマーカーとして働きうる10クラスのcDNAクローンを特定した(例えば、1762, intestinal protein)。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)。

20

30

【0012】

明確に、結腸癌、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。所与のケースの結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。結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。

【0013】

哺乳類の腸タンパク質をコードするcDNAの発見は、とりわけ結腸癌、クーロン病、及び潰瘍性大腸炎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。

40

【発明の開示】

【発明の効果】

【0014】

本発明は、特に結腸癌、クーロン病、及び潰瘍性大腸炎のような結腸疾患の診断・治療に有用な哺乳類の腸タンパク質(IPs)をコードするcDNAの発見に基づいている。

【0015】

本発明は、SEQ ID NO:1若しくはSEQ ID NO:2のアミノ酸配列、SEQ ID NO:1若しくはSEQ ID NO:2のアミノ酸配列と少なくとも80%の同一性を有する変異体、SEQ ID NO:1若しく

50

はSEQ ID NO:2の抗原性エピトープ、SEQ ID NO:1若しくはSEQ ID NO:2のオリゴペプチド、及びSEQ ID NO:1若しくはSEQ ID NO:2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。

**【0016】**

本発明はまた、SEQ ID NO:3及びSEQ ID NO:10の核酸配列、SEQ ID NO:3若しくはSEQ ID NO:10の核酸配列と少なくとも80%の同一性を有する変異体、SEQ ID NO:4乃至9を含むSEQ ID NO:3若しくはSEQ ID NO:11乃至15を含むSEQ ID NO:10の断片、及びSEQ ID NO:3乃至29のオリゴヌクレオチド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、IP-1若しくはIP-2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IP-1若しくはIP-2を作製する方法を提供する。更に本発明は、IP-1若しくはIP-2をコードするcDNAを含むベクターを有する遺伝子組換え細胞系または遺伝子組換え生物を提供する。本発明はまた、SEQ ID NO:16-29から成る群から選択される哺乳動物断片またはその相補体を提供する。或る実施態様では、本発明は少なくとも一つのそれら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。

10

**【0017】**

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、特に結腸癌、クローン病、潰瘍性大腸炎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。

20

**【0018】**

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、アプタマー、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。本発明はまた、SEQ ID NO:1若しくはSEQ ID NO:2のアミノ酸配列と、SEQ ID NO:1若しくはSEQ ID NO:2のアミノ酸配列と80%の同一性を有する変異体と、SEQ ID NO:1若しくはSEQ ID NO:2の抗原エピトープと、SEQ ID NO:1若しくはSEQ ID NO:2のオリゴペプチドと、SEQ ID NO:1若しくはSEQ ID NO:2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ほ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、とりわけ結腸癌、クローン病、及び潰瘍性大腸炎のような結腸疾患を有する被験者を治療するためにIP-1若しくはIP-2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

30

40

50

び薬剤から選択される。別の実施形態では、とりわけ結腸癌、クローン病、及び潰瘍性大腸炎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。

【0019】

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結腸癌、クローン病、及び潰瘍性大腸炎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。

10

【0020】

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。

【0021】

本発明は、とりわけ結腸癌、クローン病、及び潰瘍性大腸炎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、とりわけ結腸癌、クローン病、及び潰瘍性大腸炎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、とりわけ結腸癌、クローン病、及び潰瘍性大腸炎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結腸癌、クローン病、及び潰瘍性大腸炎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。

20

【0022】

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:3-29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。

30

【発明を実施するための最良の形態】

【0023】

(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。

40

【0024】

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

50

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。

【0025】

(定義)

「IP」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。

【0026】

「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。

10

【0027】

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。

【0028】

「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/または非移動配列(nonmoving sequence)、ゲノムDNA分子由来のイントロンを有するまたは有しないエキソンを有する。

【0029】

「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)。

20

【0030】

「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。

30

【0031】

「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。

【0032】

「障害」とは、特に結腸癌、クローン病、及び潰瘍性大腸炎である結腸疾患のような、cDNA及びIP-1若しくはIP-2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。

40

【0033】

「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。

【0034】

「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

50

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。

【0035】

「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。

【0036】

「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。

10

【0037】

「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。

【0038】

タンパク質の「翻訳後修飾」には、脂質化、グリコシル化、リン酸化、アセチル化、ラセミ化、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。

【0039】

「プローブ」とは、サンプル中の少なくとも1つの核酸にハイブリダイズするcDNAを指す。標的が一本鎖である場合、プローブは相補的一本鎖である。プローブは、サザン法、ノーザン法、in situ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。

20

【0040】

「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。

30

【0041】

「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。

【0042】

「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。

40

【0043】

「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かつ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。

【0044】

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

50

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。

【0045】

「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。

【0046】

「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。

10

【0047】

(発明)

本発明は、腸タンパク質をコードするcDNAの発見と、腎臓の疾患の性質決定、診断、治療における、cDNAまたはその断片、およびタンパク質またはその部分をそのまま使用するか組成物として使用することに基づくものである。

20

【0048】

本発明のIP-1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、結腸cDNAライブラリ(COTRNOT01)よりインサイトクローン3229449内で最初に同定された。コンセンサス配列、SEQ ID NO:3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:4-9) 2771041H1 (COLANOT02)、71851705V1、70255975V1 (SG0000124)、5596934H1 (COLCDIT03)、3229449F6 (COTRNOT01)、7128544H1 (COLNDIY01)及びGENSCAN配列GNN.g9965027-000007-006 (SEQ ID NO:30)に由来した。配列GNN.g9965027-000007-006の為に、コーディング領域はゲノミックDNAのGenscan解析によって予測された。g9965027は、Genscanが適用されたGenBank同定番号である。Genscan-予測コーディング配列GNN.g9965027-000007-006は、アセンブリに先立ってエディットされた。表1は、組織カテゴリーにかかるとIP-1転写の発現を示すものである(実施例8にリストされている)。IP-1は、消化系にて優勢に発現した。表2は、結腸癌の患者を含めて、結腸組織内の転写の発現を示す。表5は、マイクロアレイ解析で測定された通常の結腸組織に関連して、結腸癌や結腸ポリープにおけるIP-1の異なった発現を示す。IP-1は、同一患者より採取した顕微鏡学的に通常の組織に関連して、結腸癌を患った患者より採取した組織内で軽減された発現を示す。それ故に、cDNAは結腸疾患、とりわけアデノカルチノーマ、サルコーマ、クーロン病、及び潰瘍性大腸炎のような結腸癌の診断アッセイにおいて有用である。およそ170から220のヌクレオチド、およそ1015から1055のヌクレオチド、及びおよそ1500から1550のヌクレオチドよりのcDNA断片が診断アッセイに有用である。

30

40

【0049】

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。IP-1は、長さが475アミノ酸である。また、IP-1は、N29、N38、N47、N48、N92、N160、及びN210に存在する7つのN-グリコシレーション部位、S265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、T66、T225、S268、S273、S340、T354、及びS440に存在する7つの潜在的カゼインキナーゼIIリン酸化部位、S30、S49、S61、S152、S193、T298、S307、及びS340に存在する8つの潜在的プロテインキナーゼリン酸化部位、Y242及びY424に存在する2つの潜在的チロシンキナーゼリン酸化部位、及びR1113に存在する一つのRGD細胞アタッチメント配列を有する。シグナルペプチド配列はM5からQ28まで予測される。膜内外領域は、M5からV33、及びY163からL179に予測され

50

る。図3A、3B、3C、及び3Dに示されているように、IP-1は、ウサギ腸タンパク質(g1762; SEQ ID NO:32)と化学的、構造的類似性を有する。とりわけ、IP-1と腸タンパク質とは約64%の同一性、予測シグナルペプチド、2つの膜内外領域、一つの潜在的環状AMP-若しくはGMP-依存タンパク質キナーゼリン酸化部位、4つの潜在的カゼインキナーゼIIリン酸化部位、6つの潜在的プロテインキナーゼCリン酸化部位、及び2つの潜在的チロシンキナーゼリン酸化部位を持つ。有用な抗原性エピトープは、T66からH90、L120からR141、D234からR245、及びI351からN359へと広がっている。また、隣接ホモログよりIP-1を識別するのに有用なオリゴペプチドは、M1からS20およびS307からN323へと広がっている。IP-1へ特異的に結合する抗体は、とりわけ結腸癌、クーロン病、及び潰瘍性大腸炎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。

10

#### 【0050】

本発明のIP-2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、結腸cDNAライブラリ(COTRNOT13)よりインサイトクローン7484349内で最初に同定された。コンセンサス配列、SEQ ID NO:10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:11-15) 1333949F6 (COLNNOT13)、7604658J1、70106729V1 (SG0000032)、70107804V1 (SG0000032)、5865314H1 (COLTDIT04)、及びGENSCAN配列GNN.g9795680-006.edit(SEQ ID NO:31)に由来した。配列GNN.g9795680-006.editの為に、コーディング領域はゲノミックDNAのGenscan解析によって予測された。Genscan予測コーディング配列GNN.g9795680-006.editは、アセンブリに先立ってエディットされた。表3は、消化システムにおいて優勢な組織カテゴリーにかかるIP-2転写の発現を示す。表4は、結腸疾患の患者を含めて、結腸内の転写の発現を示す。表6は、マイクロアレイ解析で測定された通常の結腸組織に関連して、結腸癌や結腸ポリープにおけるIP-2の異なった発現を示す。IP-2は、同一患者より採取した顕微鏡学的に通常の組織に関連して、結腸癌を患った患者より採取した組織内で軽減された発現を示す。それ故に、cDNAは結腸疾患、とりわけアデノカルチノーマ、サルコーマ、クーロン病、及び潰瘍性大腸炎のような結腸癌の診断アッセイにおいて有用である。およそ160から210のヌクレオチド、及びおよそ1825から1875のヌクレオチドよりのcDNA断片が診断アッセイに有用である。

20

#### 【0051】

ある実施例では、本発明はSEQ ID NO: 2のアミノ酸配列を含むポリヌクレオチドを包含する。IP-2は、長さが547アミノ酸である。また、IP-2は、N30、N39、N48、N93、N161、およびN211に存在する6つのN-グリコシレーション部位、S402に存在する一つの潜在的環状AMP-若しくは環状GMP-依存タンパク質キナーゼリン酸化部位、T67、S226、T269、およびS412存在する4つの潜在的カゼインキナーゼIIリン酸化部位、S50、S153、T259、T269、T337、T356、T370、およびT487

30

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、Y243、Y267、およびY496に存在する3つの潜在的チロシンキナーゼリン酸化部位、及びR1114に存在する一つのRGD細胞アタッチメント配列を有する。膜内外領域は、L7からW35まで、及びF168からR194までに予測される。図3A、3B、3C、及び3Dに示されているように、IP-2は、ウサギ腸タンパク質(g1762; SEQ ID NO:32)と化学的、構造的類似性を有する。とりわけ、IP-2と腸タンパク質とは約78%の同一性、予測シグナルペプチド、2つの膜内外領域、一つの潜在的環状AMP-若しくはGMP-依存タンパク質キナーゼリン酸化部位、4つの潜在的カゼインキナーゼIIリン酸化部位、7つの潜在的プロテインキナーゼCリン酸化部位、及び3つの潜在的チロシンキナーゼリン酸化部位を持つ。有用な抗原性エピトープは、T104からR140まで、及びR194からK206までに広がっている。また隣接ホモログよりIP-2を識別するのに有用なオリゴペプチドは、M1からM23まで、及びR300からP313までに広がっている。IP-2へ特異的に結合する抗体は、結腸癌、クーロン病、及び潰瘍性大腸炎のような疾患に対するアッセイにおいて有用である

40

加えて、IP-1及びIP-2は、約71%の同一性を有し、予測シグナルペプチド、膜内外領域、RGD細胞アタッチメント配列、4つの潜在的N-グリコシレーション部位、一つの潜在的環状AMP-若しくはGMP-依存タンパク質キナーゼリン酸化部位、4つの潜在的カゼインキナ

50

ーゼリン酸化部位、4つの潜在的プロテインキナーゼCリン酸化部位、および2つの潜在的チロシンキナーゼリン酸化部位が共通である。

【0052】

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

【0053】

IP-1若しくはIP-2をコードするcDNAの哺乳動物変異体は、デフォルトのパラメータ及びZOOSEQデータベース(Incyte Genomics)と共にBLAST2を用いて同定された。これら好適な変異体は、以下の表に示すように約80%から約100%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。

SEQ ID <sub>H</sub>	SEQ ID <sub>var</sub>	クローン <sub>var</sub>	ライブラリ名	N <sub>H</sub> 配列	同一性
1	16	701244557H1	RALINOH02	602-807	87%
2				595-752	90%
1	17	700306567H1	RALINOT01	1220-1399	86%
1	18	700141983H1	RALINOT01	1220-1399	86%
2				357-521	84%
1	19	701725590H1	RALIUNT15	1220-1389	86%
1	20	700363665H1	RALMNON04	1250-1391,	87%,
				1469-1530	91%
1	21	701473585H1	RALITXT41	1220-1365,	86%,
				1103-1149	87%
1	22	700600759H1	RASGNOT02	687-767,	93%,
				884-923,	92%,
				819-837	100%
2				680-838,	84%,
				893-930	97%
1	23	701460109H1	RALITXT31	364-529	83%
1	24	701420417H1	RASLNOT01	960-1061	86%
1	25	701634496H1	RALIUNT13	1377-1593	80%
1	26	701601584H1	RALITXT13	1469-1593	84%
1	27	701940254H1	RALITXT1K	986-1061	86%
1	28	701463630H1	RALITXT3L	1220-1291	87%
1	29	701623610H1	RALITXT83	1495-1593	82%

【0054】

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。

【0055】

遺伝的コード悪化の結果として、IP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IP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。

【0056】

cDNA及びその断片自身(SEQ ID NOs:3-29)は、SEQ ID NO: 1、SEQ ID NO:2、およびサンプル内の関連分子間で同定及び識別可能する合成、増幅、及びスクリーニング技術において用いられても良い。ほ乳類のcDNAは、遺伝子組み換え細胞株若しくは生物を生成するのに

用いられ得る。それらは、ヒトの結腸疾患とりわけ結腸癌、クローン病、及び潰瘍性大腸炎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。

#### 【0057】

(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, 前出) 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。

#### 【0058】

##### シークエンシング

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

#### 【0059】

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

#### 【0060】

##### 核酸配列の伸長

本発明の配列は、当分野で周知の様々なPCR法を用いた方法で伸長することができる。例えば、XL-PCRキット (Applied Biosystems) 及び入れ子プライマー (nested primer)、市販のcDNAまたはゲノムDNAライブラリ用いてヌクレオチド配列を伸長することが可能である。全てのPCR系の方法に用いることができるように、プライマーは、OLIGO 4.06プライマー分析ソフトウェア (Molecular Biology Insights, Cascade CO) 等の市販のソフトウ

10

20

30

40

50

エアを用いて、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。

#### 【0061】

##### 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、IP、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3乃至29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。

10

#### 【0062】

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

20

30

#### 【0063】

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ー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)。

40

#### 【0064】

ハイブリダイゼーションプローブはまた、天然のゲノム配列のマッピングに有用である。このプローブは、(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

50

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。

【0065】

発現

IPをコードする複数のcDNAの何れか一つをベクターにクローニングして、このタンパク質若しくはその一部を宿主細胞で発現させることができる。この核酸配列を、DNAシャフリング (Stemmer and Cramer (1996) 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。

10

【0066】

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

20

【0067】

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

30

【0068】

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。宿主細胞が目的の哺乳動物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。

40

【0069】

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、

50

脂質化、及びアシル化等が含まれる。「プレプロ」型を切断する翻訳後プロセッシングを利用して、タンパク質のターゲティング、折り畳み及び/または活性を特定することができる。翻訳後活性のための特定の細胞装置及び特徴的な機構を有するATCC (Manassas, MD) から得られる異なった宿主細胞が、組み換えタンパク質の適当な修飾及びプロセッシングが確実に行われるようにするために選択され得る。

#### 【0070】

##### 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。

10

#### 【0071】

##### ペプチドの化学合成

タンパク質若しくはその一部は、組み換え方法以外の当分野で周知の化学的方法によって合成することもできる。固相技術を用いるペプチド合成は、バッチ式或いは連続的なフロープロセスによって行うことができる。連続的なフロープロセスでは、アミノ保護及び側鎖保護アミノ酸残基をリンカーを介して不溶性の高分子支持物に連続的に追加する。メチルアミン誘導体化ポリエチレングリコールなどのリンカーを、ポリ(スチレン-co-ジビニルベンゼン)に結合させて支持レジンを形成する。このアミノ酸残基は、酸不安定Boc(*t*-butyloxycarbonyl)法若しくは塩基不安定Fmoc(9-fluorenylmethoxycarbonyl)法によって保護されたN-である。保護されたアミノ酸のカルボキシル基をリンカーのアミンに結合して、この残基を固相支持レジんに結合させる。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)。

20

30

#### 【0072】

##### 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DSCR1L1 タンパク質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。

40

#### 【0073】

50

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）。

#### 【0074】

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')<sub>2</sub>断片及びこのF(ab')<sub>2</sub>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）。

10

#### 【0075】

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

20

#### 【0076】

##### 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、<sup>32</sup>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<sup>35</sup>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。

30

#### 【0077】

##### （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、結腸癌、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。

40

#### 【0078】

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。

50

## 【0079】

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。

## 【0080】

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。

## 【0081】

免疫学的方法

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

## 【0082】

（治療）

図3A、3B、3C、及び3Dに示されているように、化学的及び構造的な類似性が、IP-1（SEQ ID NO: 1）、IP-2（SEQ ID NO: 2）、及びウサギ腸タンパク質（g1762; SEQ ID NO: 32）の領域の間に存在する。加えて、IP-1もしくはIP-2の異なる発現は図1乃至6に記載されているように結腸疾患に関連している。IP-1及びIP-2は、結腸癌、クローン病、及び潰瘍性大腸炎に於ける際だった役割を有する。

## 【0083】

IP-1若しくはIP-2の増加発現に関する状況の治療には、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。

## 【0084】

IP-1若しくはIP-2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ーが、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、及びエンハンサー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関係する症状を治療する。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。

## 【0085】

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

10

20

30

40

50

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。

【0086】

#### 核酸を用いる遺伝子発現の調節

遺伝子の発現は、IP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。

10

【0087】

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

20

【0088】

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PN

30

【0089】

#### スクリーニング及び精製のアッセイ

本IP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。

40

【0090】

一実施例において、本発明のcDNAを複数の精製された分子または複合物と共にインキュベートすることが可能であり、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

50

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。

#### 【0091】

別の実施例では、cDNAは、当分野で公知であるアフィニティークロマトグラフィー方法を用いて分子あるいは化合物を精製するために使用される。一実施例では、cDNAは高分子樹脂またはゲル上で臭化シアン基と化学的に反応する。次にサンプルを通し、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。

10

#### 【0092】

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

#### 【0093】

好適な実施例では、IP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。例えば1つの方法では、組換え核酸で安定的に形質転換された、生きたまたは固定された原核宿主細胞で、細胞表面にペプチドを発現し局在させる細胞をスクリーニングアッセイで使用することが可能である。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。

20

30

#### 【0094】

或る実施態様において、本発明はUSPN 5,876,946に記載された非常に小さなアッセイ量と微量を使用して試験化合物高い処理能力でスクリーニングするための方法を網羅する。当該文献は特に引用することを以って本明細書の一部となす。この方法は、特異結合により多数の分子と化合物をスクリーニングするために用いられる。別の実施態様において本発明は、タンパク質を結合することができる中和抗体が、タンパク質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。

40

#### 【0095】

##### 薬理学

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

#### 【0096】

50

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、ED<sub>50</sub>（集団の50%に医薬的効果がある投与量）およびLD<sub>50</sub>（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、LD<sub>50</sub> / ED<sub>50</sub>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。

【0097】

（モデル系）

動物モデル系は生物学的検定法として用い得る。そこではヒトと類似の表現型反応を示し、また暴露条件が人体暴露と関連する哺乳動物は最も一般的なモデルであり、多くの感染体、癌、薬剤および毒性研究が、低費用、利便性、寿命、生殖可能性、豊富な参照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。

【0098】

毒性研究

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

【0099】

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

【0100】

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

【0101】

亜慢性毒性試験は、薬剤の連続投与に基づく。ラットとイヌはこれらの研究で一般的に用いられ、様々な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関して相当の証拠がある。

【0102】

1年以上の期間の慢性毒性試験により、毒性の不存在か薬剤による発癌の可能性を実証する。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。

【0103】

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。

【0104】

#### 胚性幹細胞

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

10

【0105】

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。

20

【0106】

#### ノックアウト分析

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

30

【0107】

#### ノックイン分析

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

【0108】

#### 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」

40

50

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。

【0109】

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

【実施例】

【0110】

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、ヒト結腸(COLANOT02)ライブラリの準備について記載する。

10

【0111】

1 cDNAの作製

COLANOT02cDNAライブラリは、慢性潰瘍性大腸炎の25歳白人女性より得た微視的には通常の結腸（上行性）組織より作成された。

【0112】

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(1gの組織/10 mlのTRIZOLの割合; Life Technologies)でホモジナイズして溶解させる。氷上での短いインキュベーションの後に、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。上部の水様相を新しい試験管に移して、イソプロパノールを加えてRNAを沈殿させた。RNAをRNアーゼのない水で再懸濁し、DNアーゼで処理した。RNAを酸性フェノール-クロロホルムで再抽出し、酢酸ナトリウムとエタノールで再沈殿した。OLIGOTEX精製キット(QIAGEN, Chatsworth CA)を用いて、ポリ(A+) RNAを単離した。

20

【0113】

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。

30

【0114】

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。

【0115】

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。

40

【0116】

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

50

カラム(Qiagen)を用いて精製した。

【0117】

### 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。

10

【0118】

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。

20

【0119】

### 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。

【0120】

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。

30

【0121】

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 2:94Cで15秒間、ステップ 3:60Cで1分間、ステップ 4:68Cで2分間、ステップ 5:ステップ 2、3及び4を20回繰り返す。ステップ 6:68Cで5分間、ステップ 7:4 で保管。別法では、プライマーの組、T7とSK+(Stratagene)に対して以下のパラメータで増幅を行った。ステップ 1:94 で3分間、ステップ 2:94Cで15秒間、ステップ 3:57Cで1分間、ステップ 4:68Cで2分間、ステップ 5:ステップ 2、3及び4を20回繰り返す。ステップ 6:68Cで5分間、ステップ 7:4 で保管。

40

【0122】

各ウェルのDNA濃度は、1 X TE及び0.5 µlの希釈していないPCR産物に溶解した100 µlのPI

50

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 $\mu$ lを1%アガロースミニゲル上で電気泳動法によって解析し、どの反応が配列の伸長に成功したかを決定した。

#### 【0123】

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、pUC18ベクター(APB)への再連結反応前に音波処理またはせん断した。ショットガン・シークエンシングのために、消化したヌ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pUC18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 $^{\circ}$ Cで一晩培養した。

#### 【0124】

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

#### 【0125】

### 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)生物のシークエンシングした。別法として、SmithとSmith(1992, Protein Engineering 5:35-51)で説明されているようなアルゴリズムを用いて、一次配列パターンと二次構造ギャップペナルティで処理することが可能であった。この出願で開示されているすべての配列は、少なくとも49ヌクレオチドの長さを有し、不要な塩基は1.2%に過ぎない(A、C、GまたはTよりもNが記録されている)。

#### 【0126】

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

#### 【0127】

無料で入手可能な配列比較アルゴリズムであるBLASTソフトウェア式(NCBI, Bethesda M

10

20

30

40

50

D、<http://www.ncbi.nlm.nih.gov/gorf/bl2.html>)には、核酸分子をアラインメントする「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 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。

10

## 【0128】

この出願の哺乳動物cDNAをアセンブルしたコンセンサス配列あるいは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・シークエ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。

20

## 【0129】

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、新たにシークエ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。

## 【0130】

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

40

## 【0131】

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G

50

enBank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。

【0132】

#### 6 染色体マッピング

スタンフォード・ヒトゲノムセンター (SHGC)、ホワイトヘッド・ゲノム研究所 (WIGR)、Genethon等の公的な情報源から入手可能な放射線ハイブリッド及び遺伝地図データを用いて、配列表に存在するcDNAが前もってマッピングされたかを測定した。マッピングされた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)。

10

【0133】

#### 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の混合を分画し、キャピラリー転移によりナイロン膜に転移する。別法として、cDNAを個々にベクターに連結反応させ、細菌性宿主細胞に挿入してライブラリを形成する。次に下記の方法の1つによってcDNAを基質に配置する。最初の方法では、個々のコロニーを含む細菌細胞をロボット利用して切りとり、ナイロン膜に配置する。膜を選択培地 (用いられたベクターに依存してカルベニシリン、カナマイシン、アンピシリンまたはクロラムフェニコール) を含むLB寒天に置き、37 °C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。

20

【0134】

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 °Cのオーブンで硬化させることにより、ポリマーコートされたスライドガラスを準備する。スライドガラスを処理前後で蒸留水中で広範囲にわたって洗浄する。核酸をスライドガラス上に置き、次にSTRATALINKER UV架橋剤 (Stratagene) を用いてアレイをUV照射に暴露することにより固定する。アレイを室温において0.2% SDSで洗浄し、蒸留水で3度洗浄する。リン酸緩衝生理食塩水 (PBS) (Tropix, Inc., Bedford MA) 中の0.2%カゼイン中において60 °C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。

30

40

【0135】

##### 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 °C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<sup>32</sup>P]dCTPをその試験管に加え、内容物を37 °C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

50

停止する。そしてPROBEQUANT G-50 microcolumn (APB)を用いてプローブを組み込まれていないヌクレオチドから精製する。精製したプローブを5分間100 に加熱する、そして2分間氷上で急速に冷却して、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。

#### 【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のプローブ調製

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ブを使用して、アレイベースのハイブリダイゼーションで配列表のcDNAをスクリーニングする。9  $\mu$ l TE バッファーで濃度200 ngにmRNAを希釈し、5  $\mu$ l 5x バッファー、1  $\mu$ l 0.1 M DTT、3  $\mu$ l Cy3またはCy5標識化混合液、1  $\mu$ l RNアーゼ阻害因子、1  $\mu$ l 逆転写酵素、5  $\mu$ 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 で2時間インキュベートする。次に反応混合液を20分間、85 でインキュベートする。そして2つの連続するCHROMA SPIN+TE 30 カラム(Clontech, Palo Alto CA)を利用してプローブを精製する。精製したプローブは、プローブをDEPC処理水の90  $\mu$ lに希釈させ、2  $\mu$ l 1mg/ml グリコーゲン、60  $\mu$ l 5 M 酢酸ナトリウム、300  $\mu$ 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  $\mu$ lの再懸濁バッファーで再懸濁し、5分間65 に加熱してから、十分に混合する。プローブを加熱して、以前のように混合して、氷上に保管する。以下に詳述するように、プローブを高密度アレイベースのハイブリダイゼーションにおいて使用する。

10

20

#### 【0137】

##### 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 Na<sub>2</sub>HPO<sub>4</sub>, 5 mM EDTA, pH 7)を含むハイブリダイゼーション溶液で、55 で2時間、膜を前もってハイブリダイズする。15 mlの新鮮なハイブリダイゼーション溶液中で希釈されたプローブを次に膜に加える。膜をプローブと共に55 で、16時間、ハイブリダイズする。ハイブリダイゼーション後、1mM Tris (pH 8.0), 1% Sarkosylで15分間、25 で膜を洗浄し、さらに1mM Tris (pH 8.0)で、各回15分間、25 で4回洗浄する。ハイブリダイゼーション化合物を検出するために、XOMAT-ARフィルム (Eastman Kodak, Rochester NY)を膜に一晩70 で暴露する。そして現像してから、視覚で確認する。

30

#### 【0138】

##### ポリマーコートされたスライドハイブリダイゼーション

プローブを5分間65 に加熱してから、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  $\mu$ 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  $\mu$ 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 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 で10分間洗浄し、次に0.1xSSCで、45 で10分間の洗浄を3回繰り返した後に、乾燥させた。

40

#### 【0139】

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

50

標識されたプローブの混合液をアレイエレメントにハイブリダイズし、2つの異なる標識からの発光を個々に検出可能な条件下で2つの異なるシグナルを調べる。両方の生物サンプルから由来した実質的に同数のプローブにハイブリダイズするアレイ上のエレメントは、特有の結合した蛍光を出す (Shalon W095/35505)。

#### 【0140】

ハイブリダイゼーション複合体は、Cy3の励起のためには488nm、Cy5の励起のためには632nmでスペクトル線を発生し得るInnova 70混合ガス10Wレーザー (Coherent, Santa Clara CA) を備えた顕微鏡で検出する。20×顕微鏡対物レンズ (Nikon, Melville NY) を用いて、アレイ上に励起レーザ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、20μmの解像度で対物レンズを通過してラスタースキャンする。示差ハイブリダイゼーションフォーマットで、レーザーにより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565nm、Cy5では650nm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

#### 【0141】

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

30

#### 【0142】

### 8 電子分析

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

40

#### 【0143】

図3で示すように、電子ノーザン分析を積スコア70で実施した。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、膵臓、呼吸器系、感覚器、皮膚、顎口腔系、非分類性/混合性または尿路がある。各カテゴリーでは、配列が発現したライブラリの数を数えて、そのカテゴリー内のライブラリの総数を示した。非ノーマライズしたライブラリでは、2つ以上の発現レベルが有意である。

#### 【0144】

50

## 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。

10

### 【0145】

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。

### 【0146】

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。

20

### 【0147】

#### 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。

30

### 【0148】

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のオープンで硬化させる。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、

40

50

スライド毎に約 5 nl の cDNA を分注する。

【 0 1 4 9 】

マイクロアレイには、STRATALINKER UVクロスリンカー (Stratagene) を用いてUV架橋する。マイクロアレイは、室温において 0.2 % SDS で 1 回洗浄し、蒸留水で 3 回洗浄する。非特異的な結合部位は、リン酸緩衝生理食塩水 (Tropix, Inc., Bedford MA) における 0.2 % カゼイン中で 60 分で 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。

【 0 1 5 0 】

### 1.1 サンプルの調合

#### 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 4097 は 48 歳の老人女性で、胃肉腫と診断されていた。ドナー 3649 は、86 歳の性別不詳の老人で、リンパ節への転移を伴う侵襲性の分化型アデノカルチノーマと診断されていた。ドナー 3839 は、60 歳の年齢不詳の老人で、結腸癌と診断されていた。比較は、同一ドナーから採取した健康な組織と腫瘍即ちポリープ組織とをマッチングすることで行われた。ドナー 3983 は、23 歳で性別不詳であり、腺腫様多発結腸ポリープよりのポリープ、及びリンパ節へ転移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。

【 0 1 5 1 】

### 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、哺乳動物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 IP 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。

【 0 1 5 2 】

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哺乳動物 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。

【 0 1 5 3 】

### 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 IP 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 IP のアミノ酸配列を解析し、免疫原性の高い領域を決定する。通常 C-末端付近或いは隣接する親水性領域内で見られる抗原性エピトープを選択して、合成し、抗体を生成するために用いられる。通常は、長さ約 15 残基のエピトープを、Fmoc ケミストリを用いる ABI 431A ペプチドシンセサイザ (Applied Biosystems) を用いて生成し、N-マレイミドベンゾイル-N-ヒドロキシスクシンイミドエステル (MBS) を用いた反応によって KLH (Sigma-Aldrich) に結合させて、免疫原性を高める

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

【 0 1 5 4 】

10

20

30

40

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。

【0155】

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<sup>32</sup>P-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。

【0156】

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

30

【0157】

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。哺乳動物タンパク質と物理的に相互作用するタンパク質をコードするcDNAを含むpB42AD-cDNAプラスミドを酵母細胞から単離して、特徴付ける。

40

【0158】

#### 1.7 IPアッセイ

50

腸に於けるIP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(前述)。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、IP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、ビ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。

【0159】

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

【0160】

(表の簡単な説明)

表1及び2は、LIFESEQ Goldデータベース(Incyte Genomics, Palo Alto Calif.)を用いて作成されたIP-1のノーザン分析を示すものである。表1では第1列が組織カテゴリを示す。第2列は、組織カテゴリ内のトータルのクローン数を示す。第3列は、少なくとも一つの転写物が全体のライブラリ数に対して見られるようなライブラリの数を示す。第4列は、転写の絶対クローン発生量を示す。第5列は転写の発生量のパーセンテージを示す。表2は、結腸疾患を有する患者よりの組織内IP-1発現を示すものである。第1列は、ライブラリ名をリストしている。第2列はライブラリのために配列されたクローンの数を示している。第3列はライブラリが誘導された組織について記載している。第4列は転写の絶対的な発生量を示す。第5列は、転写の発生量のパーセンテージを示す。

20

【0161】

表3及び4は、LIFESEQ Goldデータベース(Incyte Genomics)を用いて生成されたIP-2のノーザン分析を示すものである。表3では、第1列は、組織カテゴリを示す。第2列は、組織カテゴリ内のクローンの合計数、第3列は少なくとも一つの転写がライブラリの合計数に対して見られるライブラリの数の割合をしめす。第4列は、転写の絶対クローン発生量を示す。第5列は転写の絶対発生量のパーセンテージを示す。表4は結腸疾患を有する患者よりの組織内のIP-2発現を示すものである。第1列は、ライブラリ名をリストしている。第2列はライブラリのために配列されたクローンの数を示している。第3列はライブラリが誘導された組織について記載している。第4列は転写の絶対的な発生量を示す。第5列は、転写の発生量のパーセンテージを示す。

30

【0162】

表5は、結腸癌(例えばサルコーマ(SAR)及びアデノカルチノーマ(AdenoCA))及びマイクロアレイ解析で決定された通常の結腸組織に関連した結腸ポリープにおけるIP-1の異なった発現を示す。第1列は、結腸癌(1~4列)及び結腸ポリープ(5~6列)を有する患者より採取した組織サンプルのためにlog<sub>2</sub> 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に通常サンプルのための組織及びドナー(Dn)をリスとしたものである。第4列は、蛍光性赤ダイCy5でラベルされた病変細胞のための組織及びドナー(Dn)をリスとしたものである。

40

【0163】

表6は、結腸癌(例えばサルコーマ(SAR)及びアデノカルチノーマ(AdenoCA))及びマイクロアレイ解析で決定された通常の結腸組織に関連した結腸ポリープにおけるIP-2の異なった発現を示す。第1列は、結腸癌(1~4列)及び結腸ポリープ(5~6列)を有する患者より採取した組織サンプルのためにlog<sub>2</sub> DE(病変組織/微視的通常組織)として示された

50

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細胞のための組織及びドナー (Dn) をリストしたものである。

【図面の簡単な説明】

【0164】

【図 1 - A】図は、cDNA (SEQ ID NO: 3) によってコードされた IP-1 (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 - B】図は、cDNA (SEQ ID NO: 3) によってコードされた IP-1 (SEQ ID NO: 1) を示している。翻訳は、MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA.) によってなされた。

【図 1 - C】図は、cDNA (SEQ ID NO: 3) によってコードされた IP-1 (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: 3) によってコードされた IP-1 (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: 3) によってコードされた IP-1 (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 - F】図は、cDNA (SEQ ID NO: 3) によってコードされた IP-1 (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 2 - A】図は、cDNA (SEQ ID NO: 10) によってコードされた IP-2 (SEQ ID NO: 2) を示している。翻訳は、MACDNASIS PRO software (Hitachi Software Engineering) によってなされた。

【図 2 - B】図は、cDNA (SEQ ID NO: 10) によってコードされた IP-2 (SEQ ID NO: 2) を示している。翻訳は、MACDNASIS PRO software (Hitachi Software Engineering) によってなされた。

30

【図 2 - C】図は、cDNA (SEQ ID NO: 10) によってコードされた IP-2 (SEQ ID NO: 2) を示している。翻訳は、MACDNASIS PRO software (Hitachi Software Engineering) によってなされた。

【図 2 - D】図は、cDNA (SEQ ID NO: 10) によってコードされた IP-2 (SEQ ID NO: 2) を示している。翻訳は、MACDNASIS PRO software (Hitachi Software Engineering) によってなされた。

【図 2 - E】図は、cDNA (SEQ ID NO: 10) によってコードされた IP-2 (SEQ ID NO: 2) を示している。翻訳は、MACDNASIS PRO software (Hitachi Software Engineering) によってなされた。

40

【図 2 - F】図は、cDNA (SEQ ID NO: 10) によってコードされた IP-2 (SEQ ID NO: 2) を示している。翻訳は、MACDNASIS PRO software (Hitachi Software Engineering) によってなされた。

【図 2 - G】図は、cDNA (SEQ ID NO: 10) によってコードされた IP-2 (SEQ ID NO: 2) を示している。翻訳は、MACDNASIS PRO software (Hitachi Software Engineering) によってなされた。

【図 2 - H】図は、cDNA (SEQ ID NO: 10) によってコードされた IP-2 (SEQ ID NO: 2) を示している。翻訳は、MACDNASIS PRO software (Hitachi Software Engineering) によってなされた。

【図 3 - A】図は、IP-1 の配列及びドメイン (3229449; SEQ ID NO: 1)、IP-2 (7484349;

50

SEQ ID NO:2)及びウサギ腸タンパク質(gI762; SEQ ID NO:32)間の保存された化学的及び構造的類似性を証明する。配列はLASERGENEソフトウェア(DNASTAR, Madison Wis.)のMEGA LIGNプログラムを用いて作成された。

【図3 - B】図は、IP-1の配列及びドメイン(3229449; SEQ ID NO: 1)、IP-2 (7484349; SEQ ID NO:2)及びウサギ腸タンパク質(gI762; SEQ ID NO:32)間の保存された化学的及び構造的類似性を証明する。配列はLASERGENEソフトウェア(DNASTAR, Madison Wis.)のMEGA LIGNプログラムを用いて作成された。

【図3 - C】図は、IP-1の配列及びドメイン(3229449; SEQ ID NO: 1)、IP-2 (7484349; SEQ ID NO:2)及びウサギ腸タンパク質(gI762; SEQ ID NO:32)間の保存された化学的及び構造的類似性を証明する。配列はLASERGENEソフトウェア(DNASTAR, Madison Wis.)のMEGA LIGNプログラムを用いて作成された。 10

【図3 - D】図は、IP-1の配列及びドメイン(3229449; SEQ ID NO: 1)、IP-2 (7484349; SEQ ID NO:2)及びウサギ腸タンパク質(gI762; SEQ ID NO:32)間の保存された化学的及び構造的類似性を証明する。配列はLASERGENEソフトウェア(DNASTAR, Madison Wis.)のMEGA LIGNプログラムを用いて作成された。

【0165】

【表1】

表1

組織カテゴリ	クローン カウント	発見された数	絶対 発生量	発生量の パーセンテージ
心血管系	266190	0/68	0	0.0000
結合組織	144645	0/47	0	0.0000
消化器系	501101	9/148	10	0.0020
胚構造	106713	0/21	0	0.0000
内分泌系	225386	0/53	0	0.0000
外分泌腺	254635	0/64	0	0.0000
生殖 女性	427284	0/106	0	0.0000
生殖 男性	448207	1/114	1	0.0002
生殖細胞	38282	0/5	0	0.0000
血液及び免疫系	680277	0/159	0	0.0000
肝臓	109378	0/35	0	0.0000
筋骨格系	159280	0/47	0	0.0000
神経系	955753	0/198	0	0.0000
脾臓	110207	0/24	0	0.0000
呼吸器系	390086	0/93	0	0.0000
感覚器	19256	0/8	0	0.0000
皮膚	72292	0/15	0	0.0000
顎口腔系	12923	0/10	0	0.0000
非分類/混合	120926	0/13	0	0.0000
尿路	279052	0/64	0	0.0000
合計	5321883	10/1292	11	0.0002

20

30

【0166】

【表2】

表2

クローン カウント	ライブラリID	ライブラリの解説	絶対 発生量	発生量の パーセンテージ
1587	COLENOT02	colon, epithelium, 13F	1	0.0630
5120	COLANOP02	colon, 25M, WN	2	0.0391
3676	COLANOC03	colon, CUC	1	0.0272
3787	COLANOT23	colon, ulcerative colitis, 16M	1	0.0264
3812	COLANDY01	colon, normal/pseudopolyp, Crohn, 16, 26M, pool, lg/N	1	0.0262
3855	COTRNOT01	colon, transverse, Crohn's, 26M	1	0.0259
3902	COLANDIC01	colon, normal/pseudopolyp, Crohn, 16, 26M, pool, lg cDNA	1	0.0256
3859	COLNTUT16	colon tumor, adenocA, 60M	1	0.0259

10

20

30

【 0 1 6 7 】

【 表 3 】

表3

組織カテゴリ	クローン カウント	発見された数	絶対 発生量	発生量の パーセンテージ
心血管系	266190	0/68	0	0.0000
結合組織	144645	0/47	0	0.0000
消化器系	501101	6/148	8	0.0016
胚構造	106713	0/21	0	0.0000
内分泌系	225386	0/53	0	0.0000
外分泌腺	254635	0/64	0	0.0000
生殖 女性	427284	0/106	0	0.0000
生殖 男性	448207	0/114	0	0.0000
生殖細胞	38282	0/5	0	0.0000
血液及び免疫系	680277	0/159	0	0.0000
肝臓	109378	0/35	0	0.0000
筋骨格系	159280	0/47	0	0.0000
神経系	955753	0/198	0	0.0000
脾臓	110207	1/24	1	0.0009
呼吸器系	390086	0/93	0	0.0000
感覚器	19256	0/8	0	0.0000
皮膚	72292	0/15	0	0.0000
顎口腔系	12923	0/10	0	0.0000
非分類/混合	120926	0/13	0	0.0000
尿路	279062	0/64	0	0.0000
合計	5321883	7/1292	9	0.0002

40

【 0 1 6 8 】

50

【表4】

表4

ライブラリID	クロック カウント	ライブラリの解説	絶対 発生量	発生量の パーセンテージ
COLNNOF13	3218	co_lon, ascending, mw/CJC, 28M	2	0.0622
COLAUCT01	4034	co_lon, ascending, CUC, 74M	2	0.0496
COLFTVE01	3659	co_lon tumor, rectum, adenoca, mw/tubular adenoma, 50M, 5RP	1	0.0273
COLNTVT16	3859	colon tumor, adenoca, 60M	1	0.0259
COLNFR01	4772	co_lon, aw/Patau's, fetal., 20M, lg cdna	1	0.0210

10

20

30

【0169】

【表5】

表5

mean log2 DE (C <sub>3</sub> /C <sub>5</sub> )	C <sub>3</sub> %	C <sub>5</sub>
-2.4	8.69	Human, Colon, Sigmoid, NrmL, mw/SAR, Dn Human, Colon Tumor, Sigmoid, SAR, Dn4097
-1.8	11.59	Human, Colon Pool, NrmL, mw/Cancer, Dn3 Human, Colon Tumor, Cancer, Dn3647
-1.42	9.63	Human, Colon Pool, NrmL, mw/AdenoCA, Dn Human, Colon Tumor, AdenoCA, Dn3649
-0.89	16.32	Human, Colon, NrmL, mw/AdenoCA, Dn3839 Human, Colon Tumor, AdenoCA Dn3839
-0.42	2.58	Human, Colon, Mucosa, mw/AdenoCA, Dn3983 Human, Colon, Polyp, mw/AdenoCA, Dn3983
-0.18	1.49	Human, Colon, Mucosa, mw/AdenoCA, Dn3983 Human, Colon, Polyp, mw/AdenoCA, Dn3983

10

20

30

【 0 1 7 0 】

【 表 6 】

表6

mean log2 DE (C5/Cy3)	CV%	Cy3	Cy5
-1.34	3.87	Human, Colon Pool, Nrr1, mw/AdenoCA, FHuman, Colon Tumor, AdenoCA, Dn3649	Human, Colon Tumor, AdenoCA, Dn3649
-1.28	8.13	Human, Colon, Sigmoid, Nrr1, mw/SAR, FHuman, Colon Tumor, Sigmoid, SAR, Dn40	Human, Colon Tumor, Sigmoid, SAR, Dn40
-1.15	15.78	Human, Colon Pool, Nrr1, mw/Cancer, DHuman, Colon Tumor, Cancer, Dn3647	Human, Colon Tumor, Cancer, Dn3647
-0.34	7.14	Human, Colon, Nrr1, mw/AdenoCA, Dn3835Human, Colon Tumor, AdenoCA Dn3839	Human, Colon Tumor, AdenoCA Dn3839
0.3	6.16	Human, Colon, Mucosa, mw/AdenoCA, Dn3835Human, Colon, Polyp, mw/AdenoCA, Dn398	Human, Colon, Polyp, mw/AdenoCA, Dn398
0.4	0	Human, Colon, Mucosa, mw/AdenoCA, Dn3835Human, Colon, Polyp, mw/AdenoCA, Dn398	Human, Colon, Polyp, mw/AdenoCA, Dn398

10

20

30

【 1 - A 】

18 27 36 45 54  
 9 A TCC ATG CTA AMG CTA AAC AAA CTG GAA CTT ATA TCT GCA ATT TAT TTT GGT  
 63 72 81 90 99 108  
 ATA GAC AAG AGG TMT GCC AGT AGC ACA CTG GTG GCT TCA GAA GAA ATT CTC AAC  
 117 126 135 144 153 162  
 ACC TGA CTC GCC AGA GAG TCT ATG TAT GGG AAT GAA CAA TCT GTA AAC TAA AGG  
 171 180 189 198 207 216  
 ATC CTA ATC ATG AAA ATA AGT ATG ATA AAT TAT AG TCA CTA TTG GCA CTG TTS  
 M K I S M I N Y K S L L A L L  
 225 234 243 252 261 270  
 TTT ATA TTA GCC TCC TGG ATC ATT TTT ACA GTT TTC CAG AAC TCC ACA AMG GTT  
 F I L A S W I F T V Q N S T R K  
 279 288 297 306 315 324  
 TGG TCT GCT CTA AAC TTA TCC ATC TCC CAT CTC CAT TAC TGG AAC AAC TCC ACA AAG  
 W S A L N L S I S L H Y W N N S T K

【 1 - B 】

333 342 351 360 369 378  
 TTC TTA TTC CCT AAA ACA CCA CTG ATA TCA TTA AGC CCA CTA ACA GAG ACT GAA  
 S L P P K T P L I S L K P L T E T E  
 387 396 405 414 423  
 CTC AGA ATA AAG GAA ATC ATA GAG AAA CTA GAT CAG CAG ATC CCA CCC AGA CCT  
 L R I K E I I E K L D Q Q I P P R P  
 441 450 459 468 477 486  
 TTC ACC CAC GTG AAC ACC ACC AGC GCC ACA CAT ACC ACA GCC ACC ATC CTC  
 F T H V N T I T S A T H S T A T I L  
 495 504 513 522 531 540  
 AAC CCT CGA GAT ACG TAC TGC AGG GGA GAC CAG CTG CAC ATC CTC CTG GAG GTG  
 N P R D T Y C R G D Q L H I L L E V  
 549 558 567 576 585  
 AGG GAC CAC TTG GGA CCC AGG AAG CAA TAT GGC GGG GAT TTC CTG AGG GCC AGG  
 R D H L G R R K Q Y G G D F L R A R

【 1 - C 】

603 612 621 630 639 648  
 ATG TCT TCC CCA GCG CTG ATG GCA GGT GCT TCA GGA AAG GTG ACT GAC TTC AAC  
 M S S P A L M A G A S G K V T D F N  
 657 666 675 684 693 702  
 AAC GGC ACC TAC CTG GTC AGC TTC ACT CTG TTC TGG GAG GGC CAG GTC TCT CTG  
 N G T Y L V S F T L F W E G Q V S L  
 711 720 729 738 747 756  
 TCT CTG CTG ATC CAC CCC AGT GAA GGG GTG TCA GCT CTC TGG ACT GCA AGG  
 S L L L I H P S E G V S A L W S A R  
 765 774 783 792 801 810  
 AAC CAA GGC TAT GAC AGG GTG ATC TTC ACT GGC CAG TTT GTC AAT GGC ACT TCC  
 N Q G Y D R V I F T G Q F V N G T S  
 819 828 837 846 855 864  
 CAA GTC CAC TCT GAA TGT GGC CTG ATC CTA AAC ACA AAT GCT GAA TTG TGC CAG  
 Q V H S E C G L I L N T N A E L C Q

【 1 - D 】

873 882 891 900 909 918  
 TAC CTG GAC AAC AGA GAC CAA GAA GGC TTC TAC TGT GTG AGG CCT CAA CAC ATG  
 Y L D N R D Q E G F Y C V R P Q H M  
 927 936 945 954 963 972  
 CCC TGT GCT GCA CTC ACT CAC ATG TAT TCT AMG AAC AAG AAA GTT TCT TAT CTT  
 P C A A L T H M Y S R N K K V S Y L  
 981 990 999 1008 1017 1026  
 AGC AAA CAA GAA AGC CTC TTT GAA AGG TCA AAT GNG GGT GTA GAG ATY ATG  
 S K Q E K S L F E R S N V G V E I M  
 1035 1044 1053 1062 1071 1080  
 GAA AAA TTC AAT ACA ATT AGT GTC TCC AAA TGC AAC ACA CTG AAG TCA GTG GAT  
 E K F N T I S V S K C N T L K S V D  
 1089 1098 1107 1116 1125 1134  
 CTG CAT GAA TCT GGA AAA TTG CAA CAC CAG CTT GCT GTG GAT TTG GAT AGG AAC  
 L H E S G K L L Q H Q L A V D L D R N

【 1 - E 】

1143 1152 1170 1179 1188  
 ATC AAC ATC CAG TGG CAA AAA TAT TGT TAT CCC TTG ATA GGA TCA ATG ACC TAT  
 I N I Q W Q K Y C Y F L I G S M T Y  
 1197 1206 1215 1224 1233 1242  
 TCA GTC AAA GAG ATG GAG TAC CTC ACC CGG GCC AAT GAC AGA ACT GGA GGA  
 S V K E M E Y L T R A I D R T G G E  
 1251 1260 1269 1278 1287 1296  
 AAA AAT ACT GTC AAT GTT ATT TCC CTG GGC CAG GAT TTC AGA CCC TTT CCC AAT  
 K N T V I V I S L G Q H F R P F I  
 1305 1314 1323 1332 1341 1350  
 GAT GTT TTT ATC CGA AGG GCC CTC AAT GTC CAC AAA GCC APT CAG CAT TTT GTT  
 D V F I R R A L N V H K A I Q H L L  
 1359 1368 1377 1386 1395 1404  
 CTG AGA AGC CCA GAC ACT ATG GTT ATC ATC AAA ACA GAA AAC ATC AGC GAG ATG  
 L R S P D T M V I I K T E N I R E M

【 1 - F 】

1413 1422 1431 1440 1449 1458  
 TAC AAT GAT GCA GAA AGA TTT AGT GAC TTT CAT GGT TAC ATT CAA TAT CTC ATC  
 Y N D A E R F S D F H G Y I Q Y L I  
 1467 1476 1485 1494 1503 1512  
 ATA AAG GAC AAT TTC CAG GAT CTC AGT GTC AGT ATC ATT GAT GCC TGG GAT ATA  
 I K D I F Q D L S V S I I D A W D I  
 1521 1530 1539 1548 1557 1566  
 ACA AIT GCA TAT GCC ACA AAT AAT GTA CAC CCA CCT CAA CAT GTA GTC GGA AAT  
 T I A Y G T N N V H P P Q H V V G N  
 1575 1584 1593 1602 1611  
 CAG AAT AAT ATA TTA TTA AAC TAT ATT TGT TAA ATA ACA AAA AAA AAA A 3'  
 Q I N I L L N Y I C

【 2 - A 】

9 18 27 36 45 54  
 5' GCC AGT TAA GAA CTC ACA GGT ACT TTT CCA GCC ACC CAG ATP GGA GAG ATC AIT  
 63 72 81 90 99 108  
 AAA ACA GTG CAT TCT GTG CTA CTT GAC ACC TAT TGG GGT CCT GGA AGG AGG  
 117 126 135 144 153 162  
 CAA GAA TCC TGA GTG AAA CCT CGA GAA GTA TCC AAT AGG ACA TTC GTC ATC  
 M  
 171 180 189 198 207 216  
 TCC TCA AAT ACA ATG CTT CAA AAA ACG CTG ATC TTC TCT TTT TCA GTA  
 S S N T M L Q K T L L I L I S F S V  
 225 234 243 252 261 270  
 GTA ACC TGG ATG AAT TTT ATA AAT TCT CAG AAC TTC ACA AAG CTT TGG TCT GCT  
 V T W H I F I I S Q N F T K L W S A  
 279 288 297 306 315 324  
 CTA AAC TTA TCC ATC TCT GTC CAT TAC TGG AAC AAC TCC GCA AAG TCC TTA TTC  
 L N L S I S V H Y W N N S A K S L F

【 2 - B 】

333 342 351 360 369 378  
 CCT AAA ACA TCA CTG ATA CCA TTA AAG CCA CTA ACA GAG ACT GBA CTC AGA ATA  
 P K T S L I P L K P L T E T E L R I  
 387 396 405 414 423 432  
 AAG GAA ATC ATA GAG AAA CTA GAT CAG CAG ATC CCA CCC AGA CCT TTC ACC CAT  
 K E I I E K L D Q Q I P P R P F T H  
 441 450 459 468 477 486  
 GTG AAC ACC ACC AGT GCC ACA CAC AGC ACA GCC ACC ATC CTC AAC CPT CGA  
 V N T T S A T H S T A T I L N P R  
 495 504 513 522 531 540  
 GAT ACA TAC TGC AGG GGA GAC CAG CTG GAC ATC CTA CTG GAG GTG AGG GAC CAC  
 D T Y C R G D Q L D I L L E V R D H  
 549 558 567 576 585 594  
 TTG GCA CAG AGG AAG CAA TAT GGT GGG GAT TTC CTG AGC GCC AGG ATG TCC TCC  
 L G Q R K Q Y G G D F L R A R M S S

【 2 - C 】

603 612 621 630 639 648  
 CCA GCA CTG ACG GCA GGT GCT TCA CGA AAG GTG ATG GAC TTC AAC AAT GGC ACC  
 P A L T A G A S G K V M D F N N G T  
 657 666 675 684 693 702  
 TAC CTG GTC AGC TTC ACT CTG TTC TGG GAG GGC CAG GTC TCC CTG TCT CTG CTG  
 Y L V S F T L P W E G Q V S L S L L  
 711 720 729 738 747 756  
 CTC ATC CAC CCC AGT GAA GGG TGG GCT CTC TGG AGG GCA AGG AAC CAA GGC  
 L I H P S E G A S A L W R A R N Q G  
 765 774 783 792 801 810  
 TAT GAT AAA AAT AAT TTC AAA GGC AAA TTT GAT AAT GGC ACC TCT CAT GTC TTC  
 Y D K I I F K G K F V N G T S H V F  
 819 828 837 846 855 864  
 ACT GAA TGT GGC CTG ACC CTA AAC TCA AAT GCT GAA CTC TCT GAA TAT CTG GAT  
 T E C G L T L N S N A E L C E Y L D

【 2 - D 】

873 882 891 900 909 918  
 GAC AGA GAC CAA GAA GCC TTC TAT TGT ATG AAG CCT CNA CAC AAG CCC TGT GAG  
 D R D Q E A F Y C M K F Q H M P C E  
 927 936 945 954 963 972  
 GCT CTG ACC TAC ATG ACC ACC CCG AAT AGA GAG GTA TCT TAT CTT ACA GAC AAG  
 A L T Y M T T R N R E V S Y L T D K  
 981 990 999 1008 1017 1026  
 GAA AAC AGC CTT TTC CAC AGG TCC AAA GTG GGA GTT GAA ATG AAG GAT CGT  
 E N S L F H R S K V G V E M M K D R  
 1035 1044 1053 1062 1071 1080  
 AAA CAC AAT GAT GTC ACT AAT TGT AAC AAG AGA GAA AAA ATA GAA GAG ACA TCC  
 K H I D V T N C N K R E K I E E T C  
 1089 1098 1107 1116 1125 1134  
 CAA GTT GGA ATG AAG CCT CCT GTC CCT GGT TAT ACT TTA CAA GGA AAA TGG  
 Q V G M K P P V P G G Y T L Q G K W

【 2 - E 】

1143 1152 1161 1170 1179 1188  
 ATA ACA ACA TTT TGC AAC CAG GAT CAG TTA GAC ACA AAT AAG ATA AAT GGC TGT  
 I T T F C N Q V Q L D T I K I N G C  
 1197 1206 1215 1224 1233 1242  
 TTG AAA GGC AAA CTC AAT TAC CTC CTG GGA GAC TCT ACA CTA CTT CAG TGG ATC  
 L K G K L I Y L L G D S T L R Q W I  
 1251 1260 1269 1278 1287 1296  
 TAC TAC TTC CCC AAA GTT GTA AAA ACA CTG AAG TTT TTT GAT CTT CAY GAA ACT  
 Y Y F P K V V K T L K F F D L H E T  
 1305 1314 1323 1332 1341 1350  
 GGA ATC TTT AAG AAA CAT TTG CTT CTG GAT GCA GAA AGA CAC ACT CAG ATT CAA  
 G I F K K H L L L D A E R H T Q I Q  
 1359 1368 1377 1386 1395 1404  
 TGG AAA AAA CMT AGC TAT CCC TTC GTC ACT TTC CAG CTC TAC TCT CTG ATG GAT  
 W K K H S Y F F V T F Q L Y S L I D

【 2 - F 】

1413 1422 1431 1440 1449 1458  
 GAT GAT TAT ATC CCT GGG GAA ATT GAC CCG CTA TCA GGT GAC AAA AAC ACA GCC  
 H D Y I P R E I D R L S G D K N T A  
 1467 1476 1485 1494 1503 1512  
 ATC GTC ATC ACC TTT GGC CAG CAC TTT AGA CCA TTT CCC AAT GAC AAT TTT AAT  
 I V I T F G Q H F R P F P I D I F I  
 1521 1530 1539 1548 1557 1566  
 CCG AGG GCC ATC GGT GTT CAA AAG GCT AAT GAA AGA CTG TTC CTA AGA AGC CCA  
 R R A I G V Q K A I E R L F L R S P  
 1575 1584 1593 1602 1611 1620  
 GCC ACT AAA CTG AAT AAT AAG ACA GAA ABC ATC AGG GAG ATG CAC ATA GAG ACA  
 A T K V I I K T E N I R E M H I E T  
 1629 1638 1647 1656 1665 1674  
 GAG AGG TTT GGA GAC TTC CAT GGT TAT AAT CAC TAT AAT ATC ATC AAG GAT AAT  
 E R F G D F H G Y I H Y L I M K D I

【 2 - G 】

1683 1692 1701 1710 1719 1728  
TTC AAA GAC CTC AAC GTG GGC AYC ATT GAT GCC TGG GAC AYG ACC ATT GCA TAT  
F K D L N V G I I D A W D M T I A Y  
1737 1746 1755 1764 1773 1782  
GGC ACT GAC ACT ATC CAC CCA CCT GAT CAT CTC ATT GGA AAT CAG ATT AAC ATG  
G T D T I H P P D H V I G N Q I N M  
1791 1800 1809 1818 1827 1836  
TTC TTA AAC TAC ATT TGC TAA GGG ATA AAT ACT ATA CAA AAT CAC TAG GAA CCA  
F L N Y I C  
1845 1854 1863 1872 1881 1890  
ATC TCT GCA CAT AAT CCC ACA TGT ATT GTA AAG TAA GTT TTA CTC ATT TTA GGA  
1899 1908 1917 1926 1935 1944  
ACT AAG GAA AAT AAA TTT AAA AGR AYC TGT TTG GGG AGG AGC GCT ATG TBA GGA  
1953 1962 1971 1980 1989 1998  
CAA TGA CAA CTG ATA AGG GAT GCA AAA CCA AGA GAA TCA TTC ATG AAG AAT GAC  
2007 2016 2025 2034 2043 2052  
TMT ACC ATG CCT GGT TCT GAT CCT CGT TTA AAA TMT TAA AAA ACT TTT TTA AAA

【 2 - H 】

2061 2070 2079 2088 2097 2106  
GCC ATG TTA TTA AGC TGA TTT GAA AAT ATC TGT ACA AAT TCA TGA TGC TTT CTA  
2115 2124 2133 2142 2151 2160  
TTT CCA ATA TAG ATA TTT CCT AGC TCT GTC TAT TGA AAA GGC CTA GGA CCA ATG  
2169 2178 2187 2196 2205 2214  
ATA ACC GAT TAG CAA TAA TCA CTC CGA GCA CCC TAA CTG TGA TGT CTA AGA ACC  
2223 2232 2241  
CTT CCT CAA TAA AAG AAA AGA GGC ATC CTT GAA G 3'

【 3 - A 】

1 M K I S M I N Y K I S L L A L L - P I I A S W I I P T V E Q N 3229449  
1 M S S N T M L Q K T L L L L I L I S E S V V T M M L E I I S Q N 7484349  
1 N - - - - L H H N - Y L K L L C L L A A I C V L C I I S Q N 61762  
30 S T K V W S A L N L S I S L H Y W N N S T K S L F P K T F L 3229449  
31 F T K L W S A L N L S I S V H Y W N N S A K S L F P K T S L 7484349  
25 S T K I W G A L K L P N S H Y Y S N T S M I S I P K M S V 61762  
60 I S L K P L T E T E L R I K E I I E K L D O Q I P P R P F T 3229449  
61 I P L K P L T E T E L R I K E I I E K L D O Q I P P R P F T 7484349  
55 S P V K S L T E T E L R I K E I I E K L D L R L P P R P F T 61762  
90 H V N T T S A T H S T A T I L N P R D T Y C R G D D L H I I 3229449  
91 H V N T T S A T H S T A T I L N P R D T Y C R G D D L D I 7484349  
85 H V N T T S A T H S T A T I L N P Q D K Y C V G D D L D I 61762  
120 L L E V R D H L G R K K O Y G G D F L R A R M S S P A L M A 3229449  
121 L L E V R D H L G R K K O Y G G D F L R A R M S S P A L T A 7484349  
115 L L E V R D Y L G H Q K E Y G G D F L R A R M F I S P A L K A 61762  
150 G A S G K V T D F N N G T Y L V S F T L F W E G Q V S L S L 3229449  
151 G A S G K V D F N N G T Y L V S F T L F W E G Q V S L S L 7484349  
145 G A S G K V T D F N N G T Y L V S F T L F W E G Q V S L S V 61762

【 3 - B 】

180 L L I H P S E G V I S A L W I S I A R N Q G Y D R V I F T G Q F V 3229449  
181 L L I H P S E G A S A L W R A R N Q G Y D K I I F K G F V 7484349  
175 L L I H P S E G A S A L W R A R N Q G Y D R I I F K G F V 61762  
210 N G T S Q V H S E C G L I L N T N A E L C Q V L D N F D O E 3229449  
211 N G T S H V F T E C G L T L N S N A E L C E Y L D D R D O E 7484349  
205 N G T S H V F T E C S L T L N S N T H E C K Y L L N G R D O D 61762  
240 G F Y C V R P Q H M P C A J A L T H M Y S K N K K V S Y L S K 3229449  
241 A F Y C M K P Q H M P C E A L T Y M T T R N R E V S Y L T D 7484349  
235 V F Y C M K P Q H M P C E A L T H V T S R N R D I S Y L T S 61762  
270 Q E K S L F E R S N V G V E I M E K F N T T I S V S K C N - - 3229449  
271 K E N L F H R S K V G V E M M K D R K H I D V T C N K R 7484349  
265 K E K N L F H R S K V G V E I M K N Q - H I D V S Q C N K S 61762  
298 - - - - - 3229449  
301 E K I E T C Q V G M K P P V P G G Y T L Q G K W I T T F C 7484349  
294 K E V K E K C Q I G M K I P V P G G Y T L Q G R W L T T F C 61762  
298 - - - - - 3229449  
331 N Q V Q L D T I K I N G C L K G K L I Y L L G D S T L R O W 7484349  
324 N Q I Q L D T A K I S G C L K G K L I Y L M G D S T L R Q W 61762

【 3 - C 】

298 -- -- -- -- -- T L K S V D L H E S G K L Q H Q J I A V D L 3229449  
 361 I Y Y F P K V V K T L K F F D L H E T G I F K K H L L L D A 7484349  
 354 I Y Y L P K V M K T L K F F D L H E T G N F K K H L L L D A g1762

319 D R N I N I Q W Q K Y C Y P L I G S M T Y S V K E M E Y L T 3229449  
 391 F R H T Q I Q W K K H S Y P F V T F Q L Y S L I D H D Y I P 7484349  
 384 E K H T Q I Q W K K H S H P F V T Y Q L F S V I D H G Y I F g1762

345 R A I D R T G C E K N T V I V I S L G O H F R P P P I D V F 3229449  
 421 R E I D R L S G D K N T A I V I T F G O H F R P P P I D I F 7484349  
 414 Q E I D R L I G D K D T V I V I T F G O H F R P P I D I F g1762

379 I R R A L N V H K A I O H L L L R S P D T M V I I K T E N I 3229449  
 451 I R R A I G V Q K A I E R L F L R S P A T K V I I K T E N I 7484349  
 444 I R R A I S V R Q A I E R L F L R S P A T K V I V I K T E N I g1762

409 R E M Y N D A E R F S D F H G Y I Q Y L I I K D I F Q D L I S 3229449  
 481 R E M H I E T E R F G D F H G Y I H Y L I M K D I F K D L N 7484349  
 474 R E M H I E A E R F G D F H G Y I Q Y L T L K D I F K D L N g1762

439 V S I I D A W D I T I A Y G T N N V H P P Q H V V G N Q I N 3229449  
 511 V G I I D A W D M T I A Y G T D T H P P D H V I G N Q I N 7484349  
 504 V S V V D A W D M T I A Y G T N N V H F P D Q V I G N Q I N g1762

【 3 - D 】

469 I L L N Y I C 3229449  
 541 M F L N Y I C 7484349  
 534 M F L N Y I C g1762

【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  
13 June 2002 (13.06.2002)

PCT

(10) International Publication Number  
WO 02/46411 A2

(51) International Patent Classification: C12N 15/12, C07K 14/47, 16/18, C12Q 1/68, G01N 33/68, A01K 67/027, A61K 39/395  
(74) Agents: STREETER, David, G. et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(21) International Application Number: PCT/US01/46767  
(22) International Filing Date: 3 December 2001 (03.12.2001)  
(25) Filing Language: English  
(26) Publication Language: English  
(30) Priority Data: 09/729,454 4 December 2000 (04.12.2000) US  
(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(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, W. [US/US]; 237 41st Street #5, Oakland, CA 94611 (US). KRASNOW, Randi [US/US]; 817 Santa Fe Avenue, Stanford, CA 94305 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (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, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

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

[Continued on next page]

(54) Title: INTESTINAL PROTEINS

9 18 27 36 45 54  
5' A TCC ATG CTA AAG GTA AAC AAA CTG CAA CTT ATA TCT GCA ATT TAT TTT GGT  
63 72 81 90 99 108  
ATA GAC AAG AGG TAT GCC AGT AGC ACA CTG GTG GCT TCA GAA GAA ATT CTC AAC  
117 126 135 144 153 162  
ACC TAG CTC GCC AGA GAG TCT ATG TAT OGG ATT GAA CAA TCT GTA AAC TAA AGG  
171 180 189 198 207 216  
ATC CTA ATC ATG AAA ATA AGT ATG ATA AAT TAT AAG TCA CTA TTG GCA CTG TTG  
H K I S M I N Y K S L L A L L  
225 234 243 252 261 270  
TTT ATA TTA GCC TCC TGG ATC ATT TTT ACA GTT TTC CAG AAC TCC ACA AAG GTT  
P I L A S W I I P T V F Q N S T K V  
279 288 297 306 315 324  
TGG TCT GCT CTA AAC TTA TCC ATC TCC CTC CAT TAC TGG AAC AAC TCC ACA AAG  
W S A L N L S I S L H Y W N N S T K

(57) Abstract: The invention provides cDNAs which encode mammalian intestinal proteins. It also provides for the use of the cDNAs, fragments, complements, and variants thereof and of the encoded protein, portions thereof and antibodies thereto for diagnosis and treatment of colon disorders, particularly colon cancer, Crohn's disease, and ulcerative colitis. The invention additionally provides expression vectors and host cells for the production of the protein and a transgenic model system.



WO 02/46411 A2

**WO 02/46411 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 02/46411

PCT/US01/46767

1

**INTESTINAL PROTEINS****TECHNICAL FIELD**

This invention relates to cDNAs which encode mammalian intestinal proteins and to the  
5 use of the cDNAs and the encoded protein in the diagnosis and treatment of colon disorders,  
particularly colon cancer, Crohn's disease, and ulcerative colitis.

**BACKGROUND OF THE INVENTION**

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

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  
20 year. Colon and rectal cancers share many environmental risk factors and both are found in  
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).

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

30 For example, it is well known that abnormal patterns of DNA methylation occur  
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  
35 significantly increased in histologically normal mucosa from patients with colon cancer or in the  
benign polyps that precede cancer. This increase continues during the progression of colonic

1

WO 02/46411

PCT/US01/46767

2

neoplasms (El-Deiry *et al.* (1991) Proc Natl Acad Sci USA 88:3470-3474). Increased DNA 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 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 progress to cancer at a mean age of 44 years. The APC gene is a part of the APC- $\beta$ -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- $\beta$ -catenin-Tcf pathway and accompanies the transition from normal colonic mucosa to metastatic carcinoma. These changes include mutation 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- $\beta$ -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 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 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 $\beta$  receptor II gene which controls cell growth. Because of the potential for irreparable 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.

2

WO 02/46411

PCT/US01/46767

3

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

5 Almost all colon cancers arise from cells in which the estrogen receptor (ER) gene has been silenced. The silencing of ER gene transcription is age related and linked to hypermethylation of the ER gene (Issa *et al.* (1994) *Nature Genetics* 7:536-540). Introduction of an exogenous ER gene into cultured colon carcinoma cells results in marked suppression of growth. The connection between loss of the ER protein in colonic epithelial cells and the consequent development of cancer has not been established.

10 During normal development, morphological changes and biochemical differentiation of the intestinal epithelium are accompanied by changes in the expression of numerous genes. The intestinal epithelium is subdivided into the crypts of Lieberkuhn and the villi. In mammals, during the transition to adulthood, the crypts deepen, the ratio of villus to crypt cells decreases, and the turnover rate of enterocytes increases. In addition, the transition to an adult diet is accompanied by changes in the levels of various enzymes, including lactase-phlorizin hydrolase, lysosomal hydrolases, sucrase-isomaltase, alkaline phosphatase, and leucine aminopeptidase. Boll *et al.* (1993; *J Biol Chem* 268:12901-12911) used a subtractive hybridization method to isolate cDNA clones for genes in rabbits that are expressed in the intestines of adults but not in babies. They identified ten classes of cDNA clones that may serve as markers of intestine development and cellular differentiation (e.g. g1762, intestinal protein). Mutations in genes that control cellular differentiation and proliferation in the intestine may be linked with colon disease (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 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.

30 The discovery of cDNAs encoding mammalian intestinal proteins satisfies a need in the art by providing compositions which are useful in the diagnosis and treatment of colon disorders, particularly colon cancer, Crohn's disease, and ulcerative colitis.

#### SUMMARY OF THE INVENTION

The invention is based on the discovery of cDNAs which encode mammalian intestinal proteins (IPs), which are useful in the diagnosis and treatment of colon disorders, particularly

3

WO 02/46411

PCT/US01/46767

4

colon cancer, Crohn's disease, and ulcerative colitis.

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 the amino acid sequences of SEQ ID NO:1 (IP-1) and SEQ ID NO:2 (IP-2), a variant having at least 80% identity  
5 to the amino acid sequences of SEQ ID NO:1 or SEQ ID NO:2, an antigenic epitope of SEQ ID NO:1 or SEQ ID NO:2, an oligopeptide of SEQ ID NO:1 or SEQ ID NO:2, and a biologically active portion of SEQ ID NO:1 or SEQ ID NO:2.

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:3 and SEQ ID  
10 NO:10, a variant having at least 80% identity to the nucleic acid sequences of SEQ ID NO:3 or SEQ ID NO:10, a fragment of SEQ ID NO:3 comprising SEQ ID NOs:4-9 or a fragment of SEQ ID NO:10 comprising SEQ ID NOs:11-15, and an oligonucleotide of SEQ ID NOs:3-29. The invention additionally provides a composition, a substrate, and a probe comprising the cDNA, or the complement of the cDNA, encoding IP-1 or IP-2. The invention further provides a vector  
15 containing the cDNA, a host cell containing the vector, and a method for using the cDNA to make IP-1 or IP-2. The invention still further provides a transgenic cell line or organism comprising the vector containing the cDNA encoding IP-1 or IP-2. The invention additionally provides a mammalian fragment or the complement thereof selected from the group consisting of SEQ ID NOs:16-29. In one aspect, the invention provides a substrate containing at least one of these  
20 fragments. In a second aspect, the invention provides a probe comprising the fragment which can be used in methods of detection, screening, and purification. In a further aspect, the probe is a single stranded complementary RNA or DNA molecule.

The invention provides a method for using a cDNA to detect the differential expression of a nucleic acid in a sample comprising hybridizing a probe to the nucleic acids, thereby forming  
25 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, Crohn's disease, and ulcerative colitis.

The invention additionally provides a method for using a cDNA or a fragment or a  
30 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  
35 aspect, the molecules or compounds are selected from aptamers, DNA molecules, RNA

4

WO 02/46411

PCT/US01/46767

5

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 the amino acid sequences of SEQ ID NO:1 or SEQ ID NO:2, a variant  
5 having at least 80% identity to the amino acid sequences of SEQ ID NO:1 or SEQ ID NO:2, an antigenic epitope of SEQ ID NO:1 or SEQ ID NO:2, an oligopeptide of SEQ ID NO:1 or SEQ ID NO:2, and a biologically active portion of SEQ ID NO:1 or SEQ ID NO:2. 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 IP-1 or IP-2 to treat a  
10 subject with colon disorders, particularly colon cancer, Crohn's disease, and ulcerative colitis comprising administering to a patient in need of such treatment the composition containing the purified protein. The invention still further provides a method for using a protein to screen a library or a plurality of molecules or compounds to identify at least one ligand, the method comprising combining the protein with the molecules or compounds under conditions to allow  
15 specific binding and detecting specific binding, thereby identifying a ligand which specifically binds the protein. In one aspect, the molecules or compounds are selected from DNA molecules, RNA molecules, peptide nucleic acids, peptides, proteins, mimetics, agonists, antagonists, antibodies, immunoglobulins, inhibitors, and drugs. In another aspect, the ligand is used to treat a subject with colon disorders, particularly colon cancer, Crohn's disease, and ulcerative colitis.

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

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

The invention provides a purified antibody which binds specifically to a protein which is expressed in colon disorders, particularly colon cancer, Crohn's disease, and ulcerative colitis. The invention also provides a method of using an antibody to diagnose colon disorders, particularly colon cancer, Crohn's disease, and ulcerative colitis comprising combining the  
35 antibody comparing the quantity of bound antibody to known standards, thereby establishing the

5

WO 02/46411

PCT/US01/46767

6

presence of colon disorders, particularly colon cancer, Crohn's disease, and ulcerative colitis. The invention further provides a method of using an antibody to treat colon disorders, particularly colon cancer, Crohn's disease, and ulcerative colitis comprising administering to a patient in need of such treatment a pharmaceutical composition comprising the purified antibody.

5 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:3-29, 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  
10 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

15 Figures 1A, 1B, 1C, 1D, 1E, and 1F show IP-1 (SEQ ID NO:1) encoded by the cDNA (SEQ ID NO:3). The translation was produced using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA).

Figures 2A, 2B, 2C, 2D, 2E, 2F, 2G, and 2H show IP-2 (SEQ ID NO:2) encoded by the cDNA (SEQ ID NO:10). The translation was produced using MACDNASIS PRO software  
20 (Hitachi Software Engineering).

Figures 3A, 3B, 3C, and 3D demonstrate the conserved chemical and structural similarities among the sequences and domains of IP-1 (3229449; SEQ ID NO:1), IP-2 (7484349; SEQ ID NO:2), and rabbit intestinal protein (g1762; SEQ ID NO:32). The alignment was produced using the MEGALIGN program of LASERGENE software (DNASTAR, Madison WI).

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

35 Tables 3 and 4 show the northern analysis for IP-2 produced using the LIFESEQ Gold

6

WO 02/46411

PCT/US01/46767

7

database (Incyte Genomics). In Table 3, the first column presents the tissue categories; the second column, the total number of clones in the tissue category; the third column, the ratio of the number of libraries in which at least one transcript was found to total number of libraries; the fourth column, the absolute clone abundance of the transcript; and the fifth column, the percent abundance of the transcript. Table 4 shows expression of IP-2 in tissues from patients with colon disorders. The first column lists the library name, the second column, the number of clones sequenced for that library; the third column, description of the tissue from which the library was derived; the fourth column, absolute clone abundance of the transcript; and the fifth column, percent abundance of the transcript.

Table 5 shows the differential expression of IP-1 in colon cancer (e.g. sarcoma (SAR) and adenocarcinoma (AdenoCA)) and colon polyps relative to normal colon tissue as determined by microarray analysis. Column 1 lists the mean differential expression (DE) values presented as  $\log_2$  DE (diseased tissue/microscopically normal tissue) for tissue samples from patients with colon cancer (rows 1-4) and colon polyps (rows 5-6). 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.

Table 6 shows the differential expression of IP-2 in colon cancer (e.g. SAR and AdenoCA) and colon polyps relative to normal colon tissue as determined by microarray analysis. Column 1 lists the mean differential expression (DE) values presented as  $\log_2$  DE (diseased tissue/microscopically normal tissue) for tissue samples from patients with colon cancer (rows 1-4) and colon polyps (rows 5-6). 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

7

WO 02/46411

PCT/US01/46767

8

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

"IP" 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 sequence, an exon with or without an intron from a genomic DNA molecule.

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.

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

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

8

WO 02/46411

PCT/US01/46767

9

"Disorder" refers to conditions, diseases or syndromes in which the cDNAs and IP-1 or IP-2 are differentially expressed such as colon disorders, particularly colon cancer, Crohn's disease, and ulcerative colitis.

"Fragment" refers to a chain of consecutive nucleotides from about 200 to about 700 base pairs in length. Fragments may be used in PCR or hybridization technologies to identify related nucleic acid molecules and in binding assays to screen for a ligand. Nucleic acids and their ligands identified in this manner are useful as therapeutics to regulate replication, transcription or 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 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 to a single stranded molecule from about 18 to about 60 nucleotides in length which may be used in hybridization or amplification technologies or in regulation of replication, transcription or translation. Substantially equivalent terms are amplicon, primer, and oligomer.

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

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

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

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

9

WO 02/46411

10

PCT/US01/46767

used as part of a fusion protein to produce an antibody.

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

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

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

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

20 alignment and to achieve a more meaningful comparison between them.

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

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

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

#### THE INVENTION

35 The invention is based on the discovery of DNAs which encode intestinal proteins and on the use of the cDNAs, or fragments thereof, and proteins, or portions thereof, directly or as

WO 02/46411

PCT/US01/46767

11

compositions in the characterization, diagnosis, and treatment of colon disorders.

Nucleic acids encoding IP-1 of the present invention were first identified in Incyte Clone 3229449 from the colon cDNA library (COTRNOT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:3, was derived from the following  
5 overlapping and/or extended nucleic acid sequences (SEQ ID NOs:4-9) 2771041H1 (COLANOT02), 71851705V1, 70255975V1 (SG0000124), 5596934H1 (COLCDIT03), 3229449F6 (COTRNOT01), 7128544H1 (COLNDTY01), and GENSCAN sequence GNN.g9965027\_000007\_006 (SEQ ID NO:30). For sequence GNN.g9965027\_000007\_006, coding regions were predicted by Genscan analysis of the genomic DNA. g9965027 is the  
10 GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequence GNN.g9965027\_000007\_006 were edited prior to assembly. Table 1 shows expression of the IP-1 transcript across the tissue categories (also listed in Example VIII). IP-1 is expressed predominantly in the digestive system. Table 2 shows expression of the transcript in colon tissues, including patients with colon disorders. Table 5 shows the differential  
15 expression of IP-1 in colon cancer and colon polyps relative to normal colon tissue as determined by microarray analysis. IP-1 shows reduced expression in tissues from patients with colon cancer relative to microscopically normal tissue from the same donors. Therefore, the cDNA is useful in diagnostic assays for colon disorders, particularly colon cancers such as adenocarcinoma and sarcoma, Crohn's disease, and ulcerative colitis. Fragments of the cDNA from about nucleotide  
20 170 to about nucleotide 220, from about nucleotide 1015 to about nucleotide 1055, and from nucleotide 1500 to 1550, are also useful in diagnostic assays.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1. IP-1 is 475 amino acids in length and has seven potential N-glycosylation sites at N29, N38, N47, N48, N92, N160, and N210; one potential cyclic AMP- or  
25 cyclic GMP-dependent protein kinase phosphorylation site at S265; seven potential casein kinase II phosphorylation sites at T66, T225, S268, S273, S340, T354, and S440; eight potential protein kinase C phosphorylation sites at S30, S49, S61, S152, S193, T298, S307, and S340; two potential tyrosine kinase phosphorylation sites at Y242 and Y424; and one RGD cell attachment sequence at R113. A signal peptide sequence is predicted from M5 to Q28. Transmembrane regions are  
30 predicted from M5 to V33 and Y163 to L179. As shown in Figures 3A, 3B, 3C, and 3D, IP-1 has chemical and structural similarity with rabbit intestinal protein (g1762; SEQ ID NO:32). In particular, IP-1 and the intestinal protein share about 64% identity, a predicted signal peptide, two transmembrane regions, one potential cyclic AMP- or cyclic GMP-dependent protein kinase phosphorylation site, four potential casein kinase II phosphorylation sites, six potential protein  
35 kinase C phosphorylation sites, and two potential tyrosine kinase phosphorylation sites. Useful

11

WO 02/46411

12

PCT/US01/46767

antigenic epitopes extend from T66 to H90, L120 to R141, D234 to R245, and I351 to N359; and oligopeptides useful for distinguishing IP-1 from the nearest homolog extend from M1 to S20 and S307 to N323. An antibody which specifically binds IP-1 is useful in assays to diagnose colon disorders, particularly colon cancer, Crohn's disease, and ulcerative colitis.

5 Nucleic acids encoding IP-2 of the present invention were first identified in Incyte Clone 7484349 from the colon cDNA library (COLNNOT13) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:10, was derived from the following overlapping and/or extended nucleic acid sequences (SEQ ID NOs:11-15) 1333949F6 (COLNNOT13), 7604658J1, 70106729V1 (SG0000032), 70107804V1 (SG0000032), 5865314H1  
10 (COLTDIT04), and GENSCAN sequence GNN.g9795680\_006.edit (SEQ ID NO:31). For sequence GNN.g9795680\_006.edit, coding regions were predicted by Genscan analysis of the genomic DNA. The Genscan-predicted coding sequence GNN.g9795680\_006.edit was edited prior to assembly. Table 3 shows expression of the IP-2 transcript across the tissue categories IP-2 is expressed predominantly in the digestive system. Table 4 shows expression of the transcript  
15 in colon tissues, including patients with colon disorders. Table 6 shows the differential expression of IP-2 in colon cancer and colon polyps relative to normal colon tissue as determined by microarray analysis. IP-2 shows reduced expression in tissues from patients with colon cancer relative to microscopically normal tissue from the same donors. Therefore, the cDNA is useful in diagnostic assays for colon disorders, particularly colon cancers, such as adenocarcinoma and  
20 sarcoma, and ulcerative colitis. Fragments of the cDNA from about nucleotide 160 to about nucleotide 210 and from about nucleotide 1825 to about nucleotide 1875 are also useful in diagnostic assays.

In another embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:2. IP-2 is 547 amino acids in length and has six potential N-  
25 glycosylation sites at N30, N39, N48, N93, N161, and N211; one potential cyclic AMP- or cyclic GMP-dependent protein kinase phosphorylation site at S402; four potential casein kinase II phosphorylation sites at T67, S226, T269, and S412; eight potential protein kinase C phosphorylation sites at S50, S153, T259, T269, T337, T356, T370, and T487; three potential tyrosine kinase phosphorylation sites at Y243, Y267, and Y496; and one RGD cell attachment  
30 sequence at R114. Transmembrane regions are predicted from L7 to W35 and F168 to R194. As shown in Figures 3A, 3B, 3C, and 3D, IP-2 has chemical and structural similarity with rabbit intestinal protein (g1762; SEQ ID NO:32). In particular, IP-2 and the intestinal protein share about 78% identity, a predicted signal peptide, two transmembrane regions, one potential cyclic AMP- or cyclic GMP-dependent protein kinase phosphorylation site, four potential casein kinase  
35 II phosphorylation sites, seven potential protein kinase C phosphorylation sites, and three potential

WO 02/46411

13

PCT/US01/46767

tyrosine kinase phosphorylation sites. Useful antigenic epitopes extend from T104 to R140 and R194 to K206; and oligopeptides useful for distinguishing IP-2 from the nearest homolog extend from M1 to M23 and R300 to P313. An antibody which specifically binds IP-2 is useful in assays to diagnose colon disorders, particularly colon cancer and ulcerative colitis.

5 In addition, IP-1 and IP-2 share about 71% identity, a predicted signal peptide, a transmembrane region, an RGD cell attachment sequence, four potential N-glycosylation sites, one potential cyclic AMP- or cyclic GMP-dependent protein kinase phosphorylation site, four potential casein kinase II phosphorylation sites, four potential protein kinase C phosphorylation sites, and two potential tyrosine kinase phosphorylation sites.

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

Mammalian variants of the cDNA encoding IP-1 or IP-2 were identified using BLAST2  
15 with default parameters and the ZOOSEQ databases (Incyte Genomics). These preferred variants have from about 80% to about 100% identity as shown in the table below. The first column shows the SEQ ID for the human cDNA (SEQ ID<sub>h</sub>); the second column, the SEQ ID for the variant cDNAs (SEQ ID<sub>var</sub>); the third column, the clone number for the variant cDNAs (Clone<sub>var</sub>); the fourth column, the library name; the fifth column, the alignment of the variant cDNA to the  
20 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.

WO 02/46411

14

PCT/US01/46767

SEQ ID <sub>n</sub>	SEQ ID <sub>seq</sub>	Clone <sub>seq</sub>	Library Name	N <sub>1</sub> Alignment	Identify
1	16	701244557H1	RALINOH02	602-807	87%
-----					
2				595-752	90%
1	17	700306567H1	RALINOT01	1220-1399	86%
1	18	700141983H1	RALINOT01	1220-1399	86%
-----					
2				357-521	84%
1	19	701725590H1	RALIUNT15	1220-1389	86%
1	20	700363665H1	RALMNON04	1250-1391,	87%,
				1469-1530	91%
1	21	701473585H1	RALITXT41	1220-1365,	86%,
				1103-1149	87%
1	22	700600759H1	RASGNOT02	687-767,	93%,
				884-923,	92%,
				819-837	100%
-----					
2				680-838,	84%,
				893-930	97%
1	23	701460109H1	RALITXT31	364-529	83%
1	24	701420417H1	RALSLNOT01	960-1061	86%
1	25	701634496H1	RALIUNT13	1377-1593	80%
1	26	701601584H1	RALITXT13	1469-1593	84%
1	27	701940254H1	RALITXT1K	986-1061	86%
1	28	701463630H1	RALITXT31	1220-1291	87%
1	29	701623610H1	RALITXT83	1495-1593	82%

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

35 The cDNA and fragments thereof (SEQ ID NOs:3-29) may be used in hybridization, amplification, and screening technologies to identify and distinguish among SEQ ID NO:1, SEQ ID NO:2, and related molecules in a sample. The mammalian cDNAs may be used to produce transgenic cell lines or organisms which are model systems for human colon disorders, particularly colon cancer, Crohn's disease, and ulcerative colitis and upon which the toxicity and efficacy of potential therapeutic treatments may be tested. Toxicology studies, clinical trials, and

WO 02/46411

15

PCT/US01/46767

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

##### 5 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.

10 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.

##### 15 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 20 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 25 Biosystems), the MEGABACE 1000 DNA sequencing system (APB), and the like. The sequences may be analyzed using a variety of algorithms well known in the art and described in Ausubel *et al.* (1997; *Short Protocols in Molecular Biology*, John Wiley & Sons, New York NY, unit 7.7) and in Meyers (1995; *Molecular Biology and Biotechnology*, Wiley VCH, New York NY, pp. 856-853).

30 Shotgun sequencing may also be used to complete the sequence of a particular cloned insert of interest. Shotgun strategy involves randomly breaking the original insert into segments of various sizes and cloning these fragments into vectors. The fragments are sequenced and reassembled using overlapping ends until the entire sequence of the original insert is known. Shotgun sequencing methods are well known in the art and use thermostable DNA polymerases, 35 heat-labile DNA polymerases, and primers chosen from representative regions flanking the

WO 02/46411

16

PCT/US01/46767

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.

#### 5 Extension of a Nucleic Acid Sequence

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

#### 15 Hybridization

The cDNA and fragments thereof can be used in hybridization technologies for various purposes. A probe may be designed or derived from unique regions such as the 5' regulatory region or from a nonconserved region (i.e., 5' or 3' of the nucleotides encoding the conserved catalytic domain of the protein) and used in protocols to identify naturally occurring molecules encoding the IP, 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:3-29. 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 AFB.

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 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,

16

WO 02/46411

17

PCT/US01/46767

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 IP may be cloned into a vector and used to express the protein, or portions thereof, in host cells. The nucleic acid sequence can be engineered by such methods as DNA shuffling (USPN 5,830,721) and site-directed mutagenesis to create new restriction sites, alter glycosylation patterns, change codon preference to increase expression in a particular host, produce splice variants, extend half-life, and the like. The expression vector may contain transcriptional and translational control elements (promoters, enhancers, specific initiation signals, and polyadenylated 3' sequence) from various sources which have been selected for their efficiency in a particular host. The vector, cDNA, and regulatory elements are combined using in vitro recombinant DNA techniques, synthetic techniques, and/or in vivo genetic recombination techniques well known in the art and described in Sambrook (supra, ch. 4, 8, 16 and 17).

17

WO 02/46411

18

PCT/US01/46767

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),  $\beta$  glucuronidase, luciferase and the like, may be propagated using culture techniques. Visible markers are also used to quantify the amount of protein expressed by the introduced genes. Verification that the host cell contains the desired mammalian cDNA is based on DNA-DNA or DNA-RNA hybridizations or PCR amplification techniques.

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

#### Recovery of Proteins from Cell Culture

Heterologous moieties engineered into a vector for ease of purification include

WO 02/46411

19

PCT/US01/46767

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  $\alpha$ -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- $\alpha$ -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 IP or any portion thereof. Adjuvants such as Freund's, mineral gels, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemacyanin (KLH), and dinitrophenol may be used to increase immunological response. The oligopeptide, peptide, or portion of protein used to induce antibodies should consist of at least about five amino acids, more preferably ten amino acids, which are identical to

WO 02/46411

PCT/US01/46767

20

a portion of the natural protein. Oligopeptides may be fused with proteins such as KLH in order to produce antibodies to the chimeric molecule.

Monoclonal antibodies may be prepared using any technique which provides for the production of antibodies by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (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)<sup>2</sup> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab)<sup>2</sup> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse *et al.* (1989) *Science* 246:1275-1281.)

The IP 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).

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

#### 35 **DIAGNOSTICS**

20

WO 02/46411

21

PCT/US01/46767

The cDNAs, fragments, oligonucleotides, complementary RNA and DNA molecules, and 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 IP may be used to quantitate the protein. Disorders associated with differential expression include colon cancer, Crohn's disease, and ulcerative colitis. The diagnostic assay may use hybridization or amplification technology to compare gene expression in a biological sample from a patient to standard samples in order to detect differential gene expression. Qualitative or quantitative methods for this comparison are well known in the art.

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

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

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

#### Immunological Methods

Detection and quantification of a protein using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell

WO 02/46411

22

PCT/US01/46767

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*.)

#### 5 THERAPEUTICS

Chemical and structural similarities exist between regions of IP-1 (SEQ ID NO:1), IP-2 (SEQ ID NO:2) and the rabbit intestinal protein (gI762; SEQ ID NO:32) as shown in Figures 3A, 3B, 3C, and 3D. In addition, differential expression of IP-1 or IP-2 is associated with colon disorders as shown in Tables 1-6. IP-1 and IP-2 clearly play roles in colon cancer, Crohn's  
10 disease, and ulcerative colitis.

In the treatment of conditions associated with increased expression of the IP-1 or IP-2, 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  
15 composition comprising an inhibitor, antagonist or antibody in conjunction with a pharmaceutical carrier may be administered to a subject to treat a condition associated with the increased expression or activity of the endogenous protein. In an additional embodiment, a vector expressing the complement of the cDNA or fragments thereof may be administered to a subject to treat the disorder.

In the treatment of conditions associated with decreased expression of the IP-1 or IP-2, 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  
20 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  
30 ligands may be administered in combination with other therapeutic agents. Selection of the agents for use in combination therapy may be made by one of ordinary skill in the art according to conventional pharmaceutical principles. A combination of therapeutic agents may act synergistically to affect treatment of a particular disorder at a lower dosage of each agent.

#### Modification of Gene Expression Using Nucleic Acids

35 Gene expression may be modified by designing complementary or antisense molecules

WO 02/46411

23

PCT/US01/46767

(DNA, RNA, or PNA) to the control, 5', 3', or other regulatory regions of the gene encoding IP. Oligonucleotides designed with reference to the transcription initiation site are preferred.

Similarly, inhibition can be achieved using triple helix base-pairing which inhibits the binding of polymerases, transcription factors, or regulatory molecules (Gee *et al.* In: Huber and Carr (1994)

5 Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary molecule may also be designed to block translation by preventing binding between ribosomes and mRNA. In one alternative, a library or plurality of cDNAs or fragments thereof may be screened to identify those which specifically bind a regulatory, nontranslated sequence.

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

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

25 Screening and Purification Assays

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

In one embodiment, the cDNA of the invention may be incubated with a plurality of  
35 purified molecules or compounds and binding activity determined by methods well known in the

WO 02/46411

24

PCT/US01/46767

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 mammalian protein or a portion thereof to purify a ligand would involve combining the protein or a portion thereof with a sample under conditions to allow specific binding, detecting specific binding between the protein and ligand, recovering the bound protein, and using an appropriate chaotropic agent to separate the protein from the purified ligand.

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

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

WO 02/46411

25

PCT/US01/46767

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

5           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  
10 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.,  
15 ED<sub>50</sub> (the dose therapeutically effective in 50% of the population) and LD<sub>50</sub> (the dose lethal to 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index, and it may be expressed as the ratio, LD<sub>50</sub>/ED<sub>50</sub>. 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.

#### 20 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,  
25 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.

#### 30 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  
35 to generate a toxicity profile and to assess potential consequences on human health following exposure to the agent.

WO 02/46411

26

PCT/US01/46767

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

#### Transgenic Animal Models

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

#### Embryonic Stem Cells

Embryonic (ES) stem cells isolated from rodent embryos retain the potential to form embryonic tissues. When ES cells are placed inside a carrier embryo, they resume normal development and contribute to tissues of the live-born animal. ES cells are the preferred cells used in the creation of experimental knockout and knockin rodent strains. Mouse ES cells, such

WO 02/46411

27

PCT/US01/46767

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 gen, 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.

#### 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

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

#### Non-Human Primate Model

The field of animal testing deals with data and methodology from basic sciences such as physiology, genetics, chemistry, pharmacology and statistics. These data are paramount in evaluating the effects of therapeutic agents on non-human primates as they can be related to human health. Monkeys are used as human surrogates in vaccine and drug evaluations, and their responses are relevant to human exposures under similar conditions. Cynomolgus and Rhesus

WO 02/46411

28

PCT/US01/46767

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

In additional embodiments, the cDNAs which encode the mammalian protein may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of cDNAs that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

#### EXAMPLES

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 colon (COLANOT02) library will be described.

##### I cDNA Library Construction

The COLANOT02 cDNA library was constructed from microscopically normal colon (ascending) tissue obtained from a 25 year-old Caucasian female with chronic ulcerative colitis.

The frozen tissue was homogenized and lysed in TRIZOL reagent (1 g tissue/10 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 SUPERSOFT plasmid system (Life Technologies) which contains a NotI primer-adaptor designed to prime the first strand cDNA synthesis at the poly(A) tail of mRNAs. Double stranded cDNA was blunted, ligated to EcoRI adaptors and digested with NotI (New England Biolabs, Beverly MA). The cDNAs were fractionated on a SEPHAROSE CL4B column (APB), and those cDNAs exceeding 400 bp were ligated into pINCY plasmid (Incyte Genomics). The plasmid pINCY was subsequently transformed into DH5 $\alpha$  competent cells (Life Technologies).

##### II Construction of pINCY Plasmid

28

WO 02/46411

PCT/US01/46767

29

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

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

The cDNAs were prepared for sequencing using the MICROLAB 2200 system (Hamilton) in combination with the DNA ENGINE thermal cyclers (MJ Research). The cDNAs were sequenced by the method of Sanger and Coulson (1975; J Mol Biol 94:441-448) using an ABI PRISM 377 sequencing system (Applied Biosystems) or the MEGABACE 1000 DNA sequencing system (APB). Most of the isolates were sequenced according to standard ABI protocols and kits (Applied Biosystems) with solution volumes of 0.25x-1.0x concentrations. In the alternative, cDNAs were sequenced using solutions and dyes from APB.

### IV Extension of cDNA Sequences

The cDNAs were extended using the cDNA clone and oligonucleotide primers. One

29

WO 02/46411

PCT/US01/46767

30

primer was synthesized to initiate 5' extension of the known fragment, and the other, to initiate 3' extension of the known fragment. The initial primers were designed using 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.

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

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

The concentration of DNA in each well was determined by dispensing 100  $\mu$ l PICOGREEN quantitation reagent (0.25% reagent in 1x TE, v/v; Molecular Probes) and 0.5  $\mu$ 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  $\mu$ l to 10  $\mu$ 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 CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC18 vector (APB). For shotgun sequences, the digested nucleotide sequences were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and the agar was digested with AGARACE enzyme (Promega). Extended clones were religated using T4 DNA ligase (New England Biolabs) into pUC18 vector (APB), treated

30

WO 02/46411

31

PCT/US01/46767

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.

5 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  
10 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

15 The cDNAs of the Sequence Listing or their deduced amino acid sequences were used to query databases such as GenBank, SwissProt, BLOCKS, and the like. These databases that contain previously identified and annotated sequences or domains were searched using BLAST or BLAST 2 (Altschul et al. *supra*; Altschul, *supra*) to produce alignments and to determine which sequences were exact matches or homologs. The alignments were to sequences of prokaryotic  
20 (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).

25 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  
30 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,  
35 Bethesda MD; <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>), includes various sequence analysis

WO 02/46411

32

PCT/US01/46767

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.

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

20 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.

25 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 30 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 35

WO 02/46411

33

PCT/US01/46767

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.

#### 15 VI Chromosome Mapping

Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Génethon are used to determine if any of the cDNAs presented in the Sequence Listing have been mapped. Any of the fragments of the cDNA encoding IP 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

##### 25 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

WO 02/46411

34

PCT/US01/46767

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  $\mu$ g.

- 5 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
- 10 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
- 15 immobilized by exposing the array to UV irradiation using a STRATALINKER UV-crosslinker (Stratagene). Arrays are then washed at room temperature in 0.2% SDS and rinsed three times in distilled water. Non-specific binding sites are blocked by incubation of arrays in 0.2% casein in phosphate buffered saline (PBS; Tropix, Bedford MA) for 30 min at 60C; then the arrays are washed in 0.2% SDS and rinsed in distilled water as before.

20 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  $\mu$ l TE buffer, denaturing by heating to 100C for five min, and briefly centrifuging. The denatured cDNA is then added to a
- 25 REDIPRIME tube (APB), gently mixed until blue color is evenly distributed, and briefly centrifuged. Five  $\mu$ l of [<sup>32</sup>P]dCTP is added to the tube, and the contents are incubated at 37C for 10 min. The labeling reaction is stopped by adding 5  $\mu$ 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
- 30 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  $\mu$ l TE
- 35 buffer and adding 5  $\mu$ l 5x buffer, 1  $\mu$ l 0.1 M DTT, 3  $\mu$ l Cy3 or Cy5 labeling mix, 1  $\mu$ l RNase

WO 02/46411

35

PCT/US01/46767

inhibitor, 1  $\mu$ l reverse transcriptase, and 5  $\mu$ 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 5 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 10 ethanol precipitated by diluting probe to 90  $\mu$ l in DEPC-treated water, adding 2  $\mu$ l 1mg/ml glycogen, 60  $\mu$ l 5 M sodium acetate, and 300  $\mu$ l 100% ethanol. The probe is centrifuged for 20 min at 20,800xg, and the pellet is resuspended in 12  $\mu$ 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.

15 Membrane-based Hybridization

Membranes are pre-hybridized in hybridization solution containing 1% Sarkosyl and 1x high phosphate buffer (0.5 M NaCl, 0.1 M Na<sub>2</sub>HPO<sub>4</sub>, 5 mM EDTA, pH 7) at 55C for two hr. The probe, diluted in 15 ml fresh hybridization solution, is then added to the membrane. The membrane is hybridized with the probe at 55C for 16 hr. Following hybridization, the membrane 20 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

25 Probe is heated to 65C for five min, centrifuged five min at 9400 rpm in a 5415C microcentrifuge (Eppendorf Scientific, Westbury NY), and then 18  $\mu$ 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  $\mu$ l of 5xSSC in a corner of the chamber. The chamber containing 30 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 35 mRNA levels in the sample. In the differential hybridization format, differential expression of a

WO 02/46411

36

PCT/US01/46767

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

5 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 Imnova 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

10 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)

15 corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. The sensitivity of the scans is calibrated using the signal intensity generated by the yeast control mRNAs added to the probe mix. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that

20 location to be correlated with a weight ratio of hybridizing species of 1:100,000.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red

25 (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using the emission spectrum for each fluorophore. A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is

30 then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS program (Incyte Genomics).

#### VIII Electronic Analysis

BLAST was used to search for identical or related molecules in the GenBank or LIFSEQ databases (Incyte Genomics). The product score for human and rat sequences was calculated as

35 follows: the BLAST score is multiplied by the % nucleotide identity and the product is divided by

WO 02/46411

37

PCT/US01/46767

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

Electronic northern analysis was performed at a product score of 70 as shown in Tables 1-4. All sequences and cDNA libraries in the LIFESEQ database were categorized by system, organ/tissue and cell type. The categories included cardiovascular system, connective tissue, digestive system, embryonic structures, endocrine system, exocrine glands, female and male genitalia, germ cells, hemic/immune system, liver, musculoskeletal system, nervous system, pancreas, respiratory system, sense organs, skin, stomatognathic system, unclassified/mixed, and the urinary tract. For each category, the number of libraries in which the sequence was expressed were counted and shown over the total number of libraries in that category. In a non-normalized library, expression levels of two or more are significant.

#### **IX Complementary Molecules**

Molecules complementary to the cDNA, from about 5 (PNA) to about 5000 bp (complement of a cDNA insert), are used to detect or inhibit gene expression. These molecules are selected using OLIGO 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 mammalian protein.

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

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

WO 02/46411

38

PCT/US01/46767

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 mammalian protein.

#### X Selection of Sequences, Microarray Preparation and Use

5 Incyte clones represent template sequences derived from the LIFESEQ GOLD assembled human sequence database (Incyte Genomics). In cases where more than one clone was available for a particular template, the 5'-most clone in the template was used on the microarray. The HUMAN GENOME GEM series 1-3 microarrays (Incyte Genomics) contain 28,626 array elements which represent 10,068 annotated clusters and 18,558 unannotated clusters. For the 10 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 15 unannotated clusters.

To construct microarrays, cDNAs were amplified from bacterial cells using primers complementary to vector sequences flanking the cDNA insert. Thirty cycles of PCR increased the initial quantity of cDNAs from 1-2 ng to a final quantity of greater than 5  $\mu$ g. Amplified cDNAs were then purified using SEPHACRYL-400 columns (APB). Purified cDNAs were immobilized 20 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 25 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/ $\mu$ 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 30 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.

WO 02/46411

39

PCT/US01/46767

**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 4097 is a 48 year-old woman, diagnosed with gastric sarcoma. 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. Donor 3839 is a 60 year-old individual, sex unknown, diagnosed with colon cancer. 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 IP**

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

Spodoptera frugiperda (Sf9) insect cells are infected with recombinant Autographica californica nuclear polyhedrosis virus (baculovirus). The polyhedrin gene is replaced with the mammalian cDNA by homologous recombination and the polyhedrin promoter drives cDNA transcription. The protein is synthesized as a fusion protein with 6xHis which enables purification as described above. Purified protein is used in the following activity and to make antibodies

**XIII Production of Antibodies**

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

Rabbits are immunized with the epitope-KLH complex in complete Freund's adjuvant.

WO 02/46411

40

PCT/US01/46767

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 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.

#### XV Screening Molecules for Specific Binding with the cDNA or Protein

The cDNA, or fragments thereof, or the protein, or portions thereof, are labeled with <sup>32</sup>P-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.

#### 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 mammalian protein of the invention. A cDNA encoding the protein is inserted into the multiple cloning site of a pLexA vector, ligated, and transformed into *E. coli*. cDNA, prepared from mRNA, is inserted into the multiple cloning site of a pB42AD vector, ligated, and transformed into *E. coli* to construct a cDNA library. The pLexA plasmid and pB42AD-cDNA library constructs are isolated from *E. coli* and used in a 2:1 ratio to co-transform competent yeast EGY48[p8op-lacZ] cells using a polyethylene glycol/lithium acetate protocol. Transformed yeast cells are plated on synthetic dropout (SD) media 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

WO 02/46411

41

PCT/US01/46767

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  $\beta$ -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  $\beta$ -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 mammalian protein, is isolated from the yeast cells and characterized.

#### XVII IP Assay

The localization of IP in the intestine is detected by fluorescence microscopy as described by Boll *et al.* (*supra*). Sections of intestinal tissue are fixed with 2.5% paraformaldehyde and 0.1% glutaraldehyde and incubated with antibodies against IP. Subcellular distributions of IP 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 02/46411

42

PCT/US01/46767

<u>Tissue Category</u>	<u>Clone Count</u>	<u>Found in</u>	<u>Abs Abund</u>	<u>Pct Abund</u>
Cardiovascular System	266190	0/68	0	0.0000
Connective Tissue	144645	0/47	0	0.0000
Digestive System	501101	9/148	10	0.0020
Embryonic Structures	106713	0/21	0	0.0000
Endocrine System	225386	0/53	0	0.0000
Exocrine Glands	254635	0/64	0	0.0000
Reproductive, Female	427284	0/106	0	0.0000
Reproductive, Male	448207	1/114	1	0.0002
Germ Cells	38282	0/5	0	0.0000
Hemic and Immune System	680277	0/159	0	0.0000
Liver	109378	0/35	0	0.0000
Musculoskeletal System	159280	0/47	0	0.0000
Nervous System	955753	0/198	0	0.0000
Pancreas	110207	0/24	0	0.0000
Respiratory System	390086	0/93	0	0.0000
Sense Organs	19256	0/8	0	0.0000
Skin	72292	0/15	0	0.0000
Stomatognathic System	12923	0/10	0	0.0000
Unclassified/Mixed	120926	0/13	0	0.0000
Urinary Tract	279062	0/64	0	0.0000
<b>Totals</b>	<b>5321883</b>	<b>10/1292</b>	<b>11</b>	<b>0.0002</b>

**TABLE 1**

42

WO 02/46411

43

PCT/US01/46767

Found in:		Clone	Library Description	Count	Abs Abund	Pct Abund
Library ID	Clone					
COLENO702	1587	colon, epithelium, 13F	1587	1	0.0630	
COLNNO702	5120	colon, 25M, WN	5120	2	0.0391	
COLNUCT03	3676	colon, CUC	3676	1	0.0272	
COLNNO723	3787	colon, ulcerative colitis, 16M	3787	1	0.0264	
COLNDIT01	3812	colon, normal/pseudopoly, Crohn, 15, 26M, pool, 1g/N	3812	1	0.0262	
COFRNO701	3855	colon, transverse, Crohn's, 26M	3855	1	0.0259	
COLNDIC01	3902	colon, normal/pseudopoly, Crohn, 15, 26M, pool, 1g cDNA	3902	1	0.0256	
COLNTUT16	3859	colon tumor, adenoca, 60M	3859	1	0.0259	

TABLE 2

43

WO 02/46411

44

PCT/US01/46767

Tissue Category	Clone Count	Found in	Abs Abund	Pct Abund
Cardiovascular System	266190	0/68	0	0.0000
Connective Tissue	144645	0/47	0	0.0000
Digestive System	501101	6/148	8	0.0016
Embryonic Structures	106713	0/21	0	0.0000
Endocrine System	225386	0/53	0	0.0000
Exocrine Glands	254635	0/64	0	0.0000
Reproductive, Female	427284	0/106	0	0.0000
Reproductive, Male	448207	0/114	0	0.0000
Germ Cells	38282	0/5	0	0.0000
Hemic and Immune System	680277	0/159	0	0.0000
Liver	109378	0/35	0	0.0000
Musculoskeletal System	159280	0/47	0	0.0000
Nervous System	955753	0/198	0	0.0000
Pancreas	110207	1/24	1	0.0009
Respiratory System	390086	0/93	0	0.0000
Sense Organs	19256	0/8	0	0.0000
Skin	72292	0/15	0	0.0000
Stomatognathic System	12923	0/10	0	0.0000
Unclassified/Mixed	120926	0/13	0	0.0000
Urinary Tract	279062	0/64	0	0.0000
Totals	5321883	7/1292	9	0.0002

TABLE 3

44

WO 02/46411

45

PCT/US01/46767

Found in:		Clone	Library Description	Abs	Pct
Library ID	Count			Abund	Abund
COLANNCT13	3218		colon, ascending, mw/CUC, 28M	2	0.0622
COLALCT01	4034		colon, ascending, CUC, 74M	2	0.0496
COLRTUE01	3659		colon tumor, rectum, adenoca, mw/tubular adenoma, 50M, 5RP	1	0.0273
COLMTU016	3859		colon tumor, adenoca, 60M	1	0.0259
COLNFEC01	4772		colon, aw/Patau's, fetal, 20wk, Ig cDNA	1	0.0210

TABLE 4

45

mean log <sub>2</sub> DE (C5/C3)	CV%	C5
-2.4	8.69	Human, Colon, Sigmoid, Nml, mw/SAR, Dr Human, Colon Tumor, Sigmoid, SAR, Dnd097
-1.8	11.59	Human, Colon Pool, Nml, mw/Cancer, Dn3 Human, Colon Tumor, Cancer, Dn3647
-1.42	9.63	Human, Colon Pool, Nml, mw/AdenoCA, Du Human, Colon Tumor, AdenoCA, Dn3649
-0.89	16.32	Human, Colon, Nml, mw/AdenoCA, Dn3839 Human, Colon Tumor, AdenoCA, Dn3839
-0.42	2.58	Human, Colon, Mucosa, mw/AdenoCA, Dn3983 Human, Colon, Polyp, mw/AdenoCA, Dn3983
-0.18	1.49	Human, Colon, Mucosa, mw/AdenoCA, Dn3983 Human, Colon, Polyp, mw/AdenoCA, Dn3983

TABLE 5

mean log2 DE (C <sub>5</sub> /C <sub>3</sub> )	CV%	Cy3	Cy5
-1.34	3.67	Human, Colon Pool, Nml, mw/AdenoCA, I Human, Colon Tumor, AdenoCA, Dn3649	
-1.28	8.13	Human, Colon, Sigmoid, Nml, mw/SAR, I Human, Colon Tumor, Sigmoid, SAR, Dn40	
-1.15	15.78	Human, Colon Pool, Nml, mw/Cancer, Dr Human, Colon Tumor, Cancer, Dn3647	
-0.34	7.14	Human, Colon, Nml, mw/AdenoCA, Dn3835 Human, Colon Tumor, AdenoCA Dn3839	
0.3	6.16	Human, Colon, Mucosa, mw/AdenoCA, Dn35 Human, Colon, Polyp, mw/AdenoCA, Dn398	
0.4	0	Human, Colon, Mucosa, mw/AdenoCA, Dn35 Human, Colon, Polyp, mw/AdenoCA, Dn398	

TABLE 6

WO 02/46411

48

PCT/US01/46767

What is claimed is:

1. An isolated mammalian cDNA or a fragment thereof encoding a mammalian protein or a portion thereof selected from:
  - a) an amino acid sequence of SEQ ID NO:1 and SEQ ID NO:2;
  - 5 b) a variant having at least 80% identity to the amino acid sequence of SEQ ID NO:1 or SEQ ID NO:2;
  - c) an antigenic epitope of SEQ ID NO:1 or SEQ ID NO:2;
  - d) an oligopeptide of SEQ ID NO:1 or SEQ ID NO:2; and
  - e) a biologically active portion of SEQ ID NO:1 or SEQ ID NO:2.
- 10 2. An isolated mammalian cDNA encoding a mammalian protein of SEQ ID NO:1 or SEQ ID NO:2.
3. An isolated mammalian cDNA or the complement thereof selected from:
  - a) a nucleic acid sequence of SEQ ID NO:3 and SEQ ID NO:10;
  - b) a fragment selected from SEQ ID NOs:4-9 and SEQ ID NOs:11-15;
  - 15 c) a variant selected from SEQ ID NOs:16-29 having at least 80% identity to the nucleic acid sequences of SEQ ID NO:3 or SEQ ID NO:10; and
  - d) an oligonucleotide of SEQ ID NOs:3-29.
4. The composition comprising the cDNA or the complement of the cDNA of claim 1.
5. A substrate comprising the cDNA or the complement of the cDNA of claim 1.
- 20 6. A probe comprising the cDNA or the complement of the cDNA of claim 1.
7. A vector comprising the cDNA of claim 1.
8. A host cell comprising the vector of claim 7.
9. A method for producing a protein, the method comprising:
  - a) culturing the host cell of claim 8 under conditions for protein expression; and
  - 25 b) recovering the protein from the host cell culture.
10. A transgenic cell line or organism comprising the vector of claim 7.
11. A method for using a cDNA to detect the differential expression of a nucleic acid in a sample comprising:
  - a) hybridizing the probe of claim 6 to the nucleic acids, thereby forming hybridization
  - 30 complexes; and
  - b) comparing hybridization complex formation with a standard, wherein the comparison indicates the differential expression of the cDNA in the sample.
12. The method of claim 11 further comprising amplifying the nucleic acids of the sample prior to hybridization.
- 35 13. The method of claim 11 wherein detection of differential expression of the cDNA is

WO 02/46411

49

PCT/US01/46767

diagnostic of colon disorders, particularly colon cancer, Crohn's disease, and ulcerative colitis.

14. A method of using a cDNA to screen a plurality of molecules or compounds, the method comprising:

- 5 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.

15. The method of claim 14 wherein the molecules or compounds are selected from DNA molecules, RNA molecules, peptide nucleic acids, artificial chromosome constructions, peptides,  
10 transcription factors, repressors, and regulatory molecules.

16. A purified mammalian protein or a portion thereof selected from:

- a) an amino acid sequence of SEQ ID NO:1 and SEQ ID NO:2;  
b) a variant having at least 80% identity to the amino acid sequence of SEQ ID NO:1 or  
SEQ ID NO:2;  
15 c) an antigenic epitope of SEQ ID NO:1 or SEQ ID NO:2;  
d) an oligopeptide of SEQ ID NO:1 or SEQ ID NO:2; and  
e) a biologically active portion of SEQ ID NO:1 or SEQ ID NO:2.

17. A composition comprising the protein of claim 16.

18. A method for using a protein to screen a plurality of molecules or compounds to identify  
20 at least one ligand, the method comprising:

- a) combining the protein of claim 16 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.

19. The method of claim 18 wherein the molecules or compounds are selected from DNA molecules, RNA molecules, peptide nucleic acids, peptides, proteins, mimetics, agonists, antagonists, antibodies, immunoglobulins, inhibitors, and drugs.

20. A purified mammalian protein comprising an amino acid sequence of SEQ ID NO:1 or  
SEQ ID NO:2.

21. A purified antibody which binds specifically to the protein of claim 20.

22. A method for preparing and purifying a polyclonal antibody with the specificity of the antibody of claim 21 comprising:

- a) immunizing a animal with a protein of SEQ ID NO:1 or SEQ ID NO:2 under conditions to elicit an antibody response;  
35 b) isolating animal antibodies;

49

WO 02/46411

50

PCT/US01/46767

- 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 polyclonal antibodies.
23. A polyclonal antibody produced by the method of claim 22.
24. A method for preparing a monoclonal antibody with the specificity of the antibody of claim 21 comprising:  
a) immunizing a animal with a protein of SEQ ID NO:1 or SEQ ID NO:2 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;  
d) culturing the hybridoma cells; and  
e) isolating monoclonal antibodies from culture.
25. A monoclonal antibody produced by the method of claim 24.
26. A method for using an antibody to immunopurify a protein comprising:  
a) attaching an antibody of claim 21 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.
27. A method for using an antibody to detect expression of a protein in a sample, the method comprising:  
a) combining the antibody of claim 21 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 protein in the sample.
28. The method of claim 27 wherein complex formation is compared with standards and is diagnostic of a colon disorder, particularly colon cancer, Crohn's disease, and ulcerative colitis.
29. A composition comprising an antibody of claim 21 and a labeling moiety.
30. A composition comprising an antibody of claim 21 and a pharmaceutical agent.
31. A method for treating a colon disorder comprising administering to a subject in need of such treatment the composition of claim 31.

WO 02/46411

1/18

PCT/US01/46767

5' 9 18 27 36 45 54  
A TCC ATG CTA AAG GTA AAC AAA CTT ATA TCT GCA ATT TAT TTT GGT  
63 72 81 90 99 108  
ATA GAC AAG AGG TAT GCC AGT AGC ACA CTG GTG GCT TCA GAA GAA ATT CTC AAC  
117 126 135 144 153 162  
ACC TAG CTC GCC AGA GAG TCT ATG TAT GGG ATT GAA CAA TCT GTA AAC TAA AGG  
171 180 189 198 207 216  
ATC CTA ATC ATG AAA ATA AGT ATG ATA AAT TAT AAG TCA CTA TTG GCA CTG TTG  
M K I S M I N Y K S L L A L L  
225 234 243 252 261 270  
TTT ATA TTA GCC TCC TGG ATC ATT TTT ACA GTT TTC CAG AAC TCC ACA AAG GTT  
F I L A S W I I F T V F Q N S T K V  
279 288 297 306 315 324  
TGG TCT GCT CTA AAC TTA TCC ATC TCC CTC CAT TAC TGG AAC AAC TCC ACA AAG  
W S A L N L S I S L H Y W N N S T K

FIGURE 1A

WO 02/46411

2/18

PCT/US01/46767

```

333   342   351   360   369   378
TCC TTA TTC CCT AAA ACA CCA CTG ATA TCA TTA AAG CCA CTA ACA GAG ACT GAA
S L F P K T P L I S L K P L T E T E
-----
387   396   405   414   423   432
CTC AGA ATA AAG GAA ATC ATA GAG AAA CTA GAT CAG ATC CCA CCC AGA CCT
L R I K E I I E K L D Q Q I P P R P
-----
441   450   459   468   477'   486
TTC ACC CAC GTG AAC ACC ACC ACC GCC ACA CAT AGC ACA GCC ACC ATC CTC
F T H V N T T T S A T H S T A T I L
-----
495   504   513   522   531   540
AAC CCT CGA GAT ACG TAC TGC AGG GGA GAC CAG CTG CAC ATC CTG CTG GAG GTG
N P R D T Y C R G D Q L H I L L E V
-----
549   558   567   576   585   594
AGG GAC CAC TTG GGA CGC AGG AAG CAA TAT GGC GGG GAT TTC CTG AGG GCC AGG
R D H L G R R R K Q Y G G D F L R A R
-----

```

FIGURE 1B

```

603 612 621 630 639 648
ATG TCT CCA GCG CTG ATG GCA GGT GCT TCA GGA AAG GTG ACT GAC TTC AAC
-----
M S S P A L M A G A S G K V T D F N
657 666 675 684 693 702
AAC GGC ACC TAC CTG GTC AGC TTC ACT CTG TTC TGG GAG GGC CAG GTC TCT CTG
-----
N G T Y L V S F T L F W E G Q V S L
711 720 729 738 747 756
TCT CTG CTG CTC ATC CAC CCC AGT GAA GGG GTG TCA GCT CTC TGG AGT GCA AGG
-----
S L L L I H P S E G V S A L W S A R
765 774 783 792 801 810
AAC CAA GGC TAT GAC AGG GTG ATC TTC ACT GGC CAG TTT GTC AAT GGC ACT TCC
-----
N Q G Y D R V I F T G Q F V N G T S
819 828 837 846 855 864
CAA GTC CAC TCT GAA TGT GGC CTG ATC CTA AAC ACA AAT GCT GAA TTG TGC CAG
-----
Q V H S E C G L I L N T N A E L C Q

```

FIGURE 1C

```

873 TAC CTG GAC AAC AGA GAC CAA GAA GGC TTC TAC TGT GTG AGG CCT CAA CAC ATG 918
Y L D N R D Q E G F Y C V R P Q H M
-----
927 GCT GCA CTC ACT CAC ATG TAT TCT AAG AAC AAG AAA GTT TCT TAT CTT 972
CCC TGT GCT GCA CTC ACT CAC ATG TAT TCT AAG AAC AAG AAA GTT TCT TAT CTT
-----
P C A A L T H M Y S K N K K V S Y L
-----
981 CAA GAA AAG AGC CTC TTT GAA AGG TCA AAT GTG GGT GTA GAG ATT ATG 1026
AGC AAA CAA GAA AAG AGC CTC TTT GAA AGG TCA AAT GTG GGT GTA GAG ATT ATG
-----
S K Q E K S L F E R S N V G V E I M
-----
1035 TTC AAT ACA ATT AGT GTC TCC AAA TGC AAC ACA CTG AAG TCA GTG GAT 1080
GAA AAA TTC AAT ACA ATT AGT GTC TCC AAA TGC AAC ACA CTG AAG TCA GTG GAT
-----
E K F N T I S V S K C N T L K S V D
-----
1089 CAT GAA TCT GGA AAA TTG CAA CAC CAG CTT GCT GTG GAT TTG GAT AGG AAC 1134
CTG CAT GAA TCT GGA AAA TTG CAA CAC CAG CTT GCT GTG GAT TTG GAT AGG AAC
-----
L H E S G K L Q H Q L A V D L D R N

```

FIGURE 1D

```

1143 1152 1161 1170 1179 1188
ATC AAC CAG TGG CAA AAA TAT TGT TAT CCC TTG ATA GGA TCA ATG ACC TAT
I N I Q W Q K Y C Y P L I G S M T Y
-----
1197 1206 1215 1224 1233 1242
TCA GTC AAA GAG ATG GAG TAC CTC ACC CGG GCC ATT GAC AGA ACT GGA GGA GAA
S V K E M E Y L T R A I D R T G G E
-----
1251 1260 1269 1278 1287 1296
AAA AAT ACT GTC ATT GTT AIT TCC CTG GGC CAG CAT TTC AGA CCC TTG CCC AIT
K N T V I V I S L G Q H F R P F I
-----
1305 1314 1323 1332 1341 1350
GAT GTT TTT ATC CGA AGG GCC CTC AAT GTC CAC AAA GCC AIT CAG CAT CTT CTT
D V F I R R A L N V H K A I Q H L L
-----
1359 1368 1377 1386 1395 1404
CTG AGA AGC CCA GAC ACT ATG GTT ATC ATC AAA ACA GAA AAC ATC AGG GAG ATG
L R S P D T M V I I K T E N I R E M
-----

```

FIGURE 1E

WO 02/46411

6/18

PCT/US01/46767

```

1413      1422      1431      1440      1449      1458
TAC AAT GAT GCA GAA AGA TTT AGT GAC TTT CAT GGT TAC ATT CAA TAT CTC ATC
-----
Y N D A E R F S D F H G Y I Q Y L I
1467      1476      1485      1494      1503      1512
ATA AAG GAC ATT TTC CAG GAT CTC AGT GTG AGT ATC ATT GAT GCC TGG GAT ATA
-----
I K D I F Q D L S V S I I D A W D I
1521      1530      1539      1548      1557      1566
ACA ATT GCA TAT GGC ACA AAT AAT GTA CAC CCA CCT CAA CAT GTA GTC GGA AAT
-----
T I A Y G T N N V H P P Q H V V G N
1575      1584      1593      1602      1611
CAG ATT AAT ATA TTA TTA AAC TAT ATT TGT TAA ATA ACA AAA AAA AAA AAA A 3'
-----
Q I N I L L N Y I C

```

FIGURE 1F

WO 02/46411

7/18

PCT/US01/46767

```

5'   9   18   27   36   45   54
    GCC AGT TAA GAA CTC ACA GGT ACT TTT CCA GCC ACC CAG ATA GGA GAG ATC ATT
    63   72   81   90   99   108
    AAA ACA GTG CAT TCT GTG CTA CCT GAC ACC TAT TGG GGT CCT GGA AGG AGG AAG
    117  126  135  144  153  162
    CAA CAA TCC TGA GTG AAA CCT CGA CAA GAA GTA TCC AAT AGG ACA TTC GTC ATG
    ---
    M

    171  180  189  198  207  216
    TCC TCA AAT ACA ATG CTT CAA AAA ACG CTG ATC TTG ATC TCT TTT TCA GTA
    ---
    S S N T M L Q K T L L I L I S F S V
    225  234  243  252  261  270
    GTA ACC TGG ATG ATT TTT ATA ATT TCT CAG AAC TTC ACA AAG CTT TGG TCT GCT
    ---
    V T W M I F I I S Q N F T K L W S A

    279  288  297  306  315  324
    CTA AAC TTA TCC ATC TCT GTC CAT TAC TGG AAC AAC TCC GCA AAG TCC TTA TTC
    ---
    L N L S I S V H Y W N N S A K S L F
  
```

FIGURE 2A

```

333      342      351      360      369      378
CCT AAA ACA TCA CTG ATA CCA TTA AAG CCA CTA ACA GAG ACT GAA CTC AGA ATA
-----
P   K   T   S   L   I   P   L   K   P   L   T   E   T   E   L   R   I
-----
387      396      405      414      423      432
AAG GAA ATC ATA GAG AAA CTA GAT CAG CAG ATC CCA CCC AGA CCT TTC ACC CAT
-----
K   E   I   I   E   K   L   D   Q   Q   I   P   P   R   P   F   T   H
-----
441      450      459      468      477      486
GTG AAC ACC ACC AGT GCC ACA CAC AGC AGC ACC ACC ATC CTC AAC CCT CGA
-----
V   N   T   T   T   S   A   T   H   S   T   A   T   I   L   N   P   R
-----
495      504      513      522      531      540
GAT ACA TAC TGC AGG GGA GAC CAG CTG GAC ATC CTA CTG GAG GTG AGG GAC CAC
-----
D   T   Y   C   R   G   D   Q   L   D   I   L   L   E   V   R   D   H
-----
549      558      567      576      585      594
TTG GGA CAG AGG AAG CAA TAT GGT GGG GAT TTC CTG AGG GCC AGG ATG TCC TCC
-----
L   G   Q   R   K   Q   Y   G   G   D   F   L   R   A   R   M   S   S

```

FIGURE 2B

```

603 612 621 630 639 648
CCA GCA CTG ACG GCA GGT GCT TCA GGA AAG GTG ATG GAC TTC AAC AAT GGC ACC
P A L T A G A S G K V M D F N N G T
657 666 675 684 693 702
TAC CTG GTC AGC TTC ACT CTG TTC TGG GAG GGC CAG GTC TCC CTG TCT CTG CTG
Y L V S F T L F W E G Q V S L S L L
711 720 729 738 747 756
CTC ATC CAC CCC AGT GAA GGG GCG TCG GCT CTC TGG AGG GCA AGG AAC CAA GGC
L I H P S E G A S A L W R A R N Q G
765 774 783 792 801 810
TAT GAT AAA ATT ATT TTC AAA GGC AAA TTT GTT AAT GGC ACC TCT CAT GTC TTC
Y D K I I F K G K F V N G T S H V F
819 828 837 846 855 864
ACT GAA TGT GGC CTG ACC CTA AAC TCA AAT GCT GAA CTC TGT GAA TAT CTG GAT
T E C G L T L N S N A E L C E Y L D

```

FIGURE 2C

```

873      882      891      900      909      918
GAC AGA GAC CAA GAA GCC TTC TAT TGT ATG AAG CCT CAA CAC ATG CCC TGT GAG
-----
D R D Q E A F Y C M K P Q H M P C E

927      936      945      954      963      972
GCT CTG ACC TAC ATG ACC ACC CGG AAT AGA GAG GTA TCT TAT CTT ACA GAC AAG
-----
A L T Y M T T R N R E V S Y L T D K

981      990      999      1008      1017      1026
GAA AAC AGC CTT TTC CAC AGG TCC AAA GTG GGA GTT GAA ATG ATG AAG GAT CGT
-----
E N S L F H R S K V G V E M M K D R

1035     1044     1053     1062     1071     1080
AAA CAC ATT GAT GTC ACT AAT TGT AAC AAG AGA GAA AAA ATA GAA GAG ACA TGC
-----
K H I D V T N C N K R E K I E E T C

1089     1098     1107     1116     1125     1134
CAA GTT GGA ATG AAG CCT CCT GTC CCT GGT GGT TAT ACT TTA CAA GGA AAA TGG
-----
Q V G M K K P P V P G G Y T L Q G K W

```

FIGURE 2D

WO 02/46411

11/18

PCT/US01/46767

```

1143      1152      1161      1170      1179      1188
ATA ACA ACA TTT TGC AAC CAG GTT CAG TTA GAC ACA ATT AAG ATA AAT GGC TGT
-----
I T T F C N Q V Q L D T I K I N G C

1197      1206      1215      1224      1233      1242
TTG AAA GGC AAA CTC ATT TAC CTC CTG GGA GAC TCT ACA CTA CGT CAG TGG ATC
-----
L K G K L I Y L L G D S T L R Q W I

1251      1260      1269      1278      1287      1296
TAC TAC TTC CCC AAA GTT GTA AAA ACA CTG AAG TTT TTT GAT CTT CAT GAA ACT
-----
Y Y F P K V V K T L K F F D L L H E T

1305      1314      1323      1332      1341      1350
GGA ATC TTT AAG AAA CAT TTG CTT CTG GAT GCA GAA AGA CAC ACT CAG ATT CAA
-----
G I F K K H L L L D A E R H T Q I Q

1359      1368      1377      1386      1395      1404
TGG AAA AAA CAT AGC TAT CCC TTC GTC ACT TTC CAG CTC TAC TCT CTG ATA GAT
-----
W K K H S Y P F V T F Q L Y S L I D

```

FIGURE 2E

```

1413      1422      1431      1440      1449      1458
CAT GAT TAT ATC CCT CGG GAA APT GAC CGG CTA TCA GGT GAC AAA AAC ACA GCC
-----
H D Y I P R E I D R L S G D K N T A
1467      1476      1485      1494      1503      1512
ATC GTC ATC ACC TTT GGC CAG CAC TTT AGA CCA TTT CCC ATT GAC ATT TTT ATT
-----
I V I T F G Q H F R P F P I D I F I
1521      1530      1539      1548      1557      1566
CGC AGG GCC ATC GGT GTT CAA AAG GCT ATT GAA AGA CTG TTC CTA AGA AGC CCA
-----
R R A I G V Q K A I E R L F L R S P
1575      1584      1593      1602      1611      1620
GCC ACT AAA GTG ATT ATT AAG ACA GAA AAC ATC AGG GAG ATG CAC ATA GAG ACA
-----
A T K V I I K T E N I R E M H I E T
1629      1638      1647      1656      1665      1674
GAG AGG TTT GGA GAC TTC CAT GGT TAT ATT CAC TAT CTT ATC ATG AAG GAT ATT
-----
E R F G G D F H G Y I H Y L I M K D I

```

FIGURE 2F

```

1683      1692      1701      1710      1719      1728
TTC AAA GAC CTC AAC GTG GGC ATC ATT GAT GCC TGG GAC ATG ACC ATT GCA TAT
-----
F K D L N V G I I D A W D M T I A Y
-----
1737      1746      1755      1764      1773      1782
GGC ACT GAC ACT ATC CAC CCA CCT GAT CAT GTG ATT GGA AAT CAG ATT AAC ATG
-----
G T D T I H P P D H V I G N Q I N M
-----
1791      1800      1809      1818      1827      1836
TTC TTA AAC TAC ATT TGC TAA GGG ATA AAT ACT ATA CAA AAT CAC TAG GAA CCA
-----
F L N Y I C
-----
1845      1854      1863      1872      1881      1890
ATC TCT GCA CAT AAT CCC ACA TGT ATT GTA AAG TAA GTT TTA CTC ATT TTA GGA
-----
1899      1908      1917      1926      1935      1944
ACT AAG GAA AAT AAA TTT AAA AGA ATC TGT TTG GGG AGG AAG GCT ATG TAA GGA
-----
1953      1962      1971      1980      1989      1998
CAA TGA CAA CTG ATA AGG GAT GCA AAA CCA AGA GAA TCA TTC ATG AAG AAT GAC
-----
2007      2016      2025      2034      2043      2052
TAT ACC ATG CCT GGT TCT GAT GCT CGT TTA AAA TAT TAA AAA AGT TTT TTA AAA

```

FIGURE 2G

2061            2070            2079            2088            2097            2106  
GCC ATG TTA TTA AGC TGA TTT GAA AAT ATC TGT ACA AAT TCA TGA TGC TTT CTA  
2115            2124            2133            2142            2151            2160  
TTT CCA ATA TAG ATA TTT CCT AGC TCT GTC TAT TGA AAA GGC CTA GGA GCA ATG  
2169            2178            2187            2196            2205            2214  
ATA ACC CAT TAG CAA TAA TCA CTC CGA GCA CCC TAA CTG TGA TGT CTA AGA ACC  
2223            2232            2241  
CTT CCT CAA TAA AAG AAA AGA GGC ATC CTT GAA G 3'

**FIGURE 2H**

WO 02/46411

15/18

PCT/US01/46767

1	M	K	I	S	M	I	N	Y	K	S	L	L	A	L	L	-	F	I	L	A	S	W	I	F	T	V	F	Q	N	3229449	
1	M	S	S	N	T	M	L	Q	K	T	L	L	I	L	I	S	E	S	V	T	W	M	I	F	I	I	S	Q	N	7484349	
1	M	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	g1762		
30	S	T	K	V	W	S	A	L	N	L	S	I	S	L	H	Y	W	N	N	S	T	K	S	L	F	P	K	T	P	L	3229449
31	F	T	K	L	W	S	A	L	N	L	S	I	S	V	H	Y	W	N	N	S	A	K	S	L	F	P	K	T	S	L	7484349
25	S	T	K	I	W	G	A	L	K	L	P	N	S	H	Y	S	N	T	S	M	I	S	I	P	K	M	S	V		g1762	
60	I	S	L	K	P	L	T	E	T	E	L	R	I	K	E	I	I	E	K	L	D	Q	Q	I	P	P	R	P	F	T	3229449
61	I	P	L	K	P	L	T	E	T	E	L	R	I	K	E	I	I	E	K	L	D	Q	Q	I	P	P	R	P	F	T	7484349
55	S	P	V	K	S	L	T	E	T	E	L	R	V	K	E	I	L	E	K	L	D	R	L	I	P	P	R	P	F	T	g1762
90	H	V	N	T	T	S	A	T	H	S	T	A	T	I	L	N	P	R	D	T	Y	C	R	G	D	Q	L	H	I	3229449	
91	H	V	N	T	T	S	A	T	H	S	T	A	T	I	L	N	P	R	D	T	Y	C	R	G	D	Q	L	D	I	7484349	
85	H	V	N	T	T	S	A	T	H	S	T	A	T	I	L	N	P	Q	D	K	Y	C	V	G	D	Q	L	D	I	g1762	
120	L	L	E	V	R	D	H	L	G	R	R	K	Q	Y	G	G	D	F	L	R	A	R	M	S	S	P	A	L	M	A	3229449
121	L	L	E	V	R	D	H	L	G	Q	R	K	Q	Y	G	G	D	F	L	R	A	R	M	S	S	P	A	L	T	A	7484349
115	L	L	E	V	R	D	Y	L	G	H	Q	K	E	Y	G	G	D	F	L	R	A	R	M	F	S	P	A	L	K	A	g1762
150	G	A	S	G	K	V	T	D	F	N	N	G	T	Y	L	V	S	F	T	L	F	W	E	G	Q	V	S	L	S	L	3229449
151	G	A	S	G	K	V	M	D	F	N	N	G	T	Y	L	V	S	F	T	L	F	W	E	G	Q	V	S	L	S	L	7484349
145	G	A	S	G	K	V	T	D	F	N	N	G	T	Y	L	V	S	F	T	L	F	W	E	G	Q	V	S	L	S	V	g1762

FIGURE 3A

180 L L I H P S E G V S A L W S A R N Q G Y D R V I F T G Q F V 3229449  
181 L L I H P S E G A S A L W R A R N Q G Y D K I I F K G K F V 7484349  
175 L L I H P S E G A S A L W R A R N Q G Y D R I I F K G Q F V g1762  
  
210 N G T S Q V H S E C G L I L N T M A E L C Q Y L D N R D Q E 3229449  
211 N G T S H V F T E C G L T L N S N A E L C E Y L D D R D Q E 7484349  
205 N G T S H V F T E C I S L T L N S N T E E C K Y L N G R D Q D g1762  
  
240 G F Y C V R P Q H M P C A A L T H M Y S K N K K V S Y L S K 3229449  
241 A F Y C M K P Q H M P C E A L T Y M T T R N R R E V S Y L T D 7484349  
235 V F Y C M K P Q H M P C E A L T H V T S R N R D I S Y L T S g1762  
  
270 Q E K S L F E R S N V G V E I M E K F N T I S V S K C N - 3229449  
271 K E N S L F H R S K V G V E M M K D R K H I D V T N C N K R 7484349  
265 K E K N L F H R S K V G V E I M K N Q - H I D V S Q C N K S g1762  
  
298 - - - - - 3229449  
301 E K I E T C Q V G M K P P V P G G Y T L Q G K W I T T F C 7484349  
294 K E V K E K C Q I G M K I P V P G G Y T L Q G R W L T T F C g1762  
  
298 - - - - - 3229449  
331 N Q V Q L D T I K I N G C L K G K L I Y L L G D S T L R Q W 7484349  
324 N Q I Q L D T A K I S G C L K G K L I Y L M G D S T L R Q W g1762

FIGURE 3B

WO 02/46411

17/18

PCT/US01/46767

298 - - - - - T L K S V D L H E S G K L Q H Q L A V D L 3229449  
 361 I Y F P K V V K T L K F F D L H E T G I F K K H L L L D A 7484349  
 354 I Y L P K V M K T L K F F D L H E T G N F K K H L L L D A g1762  
  
 319 D R N I N I Q W Q K Y C Y P L I G S M T Y S V K E M E Y L T 3229449  
 391 E R H T Q I Q W K K H S Y P F V T F Q L Y S L I D H D Y I P 7484349  
 384 E R H T Q I Q W K K H S H P F V T Y Q L F S V I D H G Y I P g1762  
  
 349 R A I D R T G G E K N T V I V I S L G Q H F R P P P I D V F 3229449  
 421 R E I D R I S G D K N T A I V I T F G Q H F R P P P I D I F 7484349  
 414 Q E I D R L I G D K D T V I V I T F G Q H F R P P P I D I F g1762  
  
 379 I R R A L N V H K A I Q H L L L R S P D T M V I I K T E N I 3229449  
 451 I R R A I G V Q K A I E R L F L R S P A T K V I I K T E N I 7484349  
 444 I R R A I S V R Q A I E R L F L R S P A T K V I V K T E N I g1762  
  
 409 R E M Y N D A E R F S D F H G Y I Q Y L I I K D I F Q D L S 3229449  
 481 R E M H I E T E R F G D F H G Y I H Y L I M K D I F K D L N 7484349  
 474 R E M H I E A E R F G D F H G Y I Q Y L T L K D I F K D L N g1762  
  
 439 V I S I I D A W D I T I A Y G T N N V H P P Q H V V I G N Q I N 3229449  
 511 V G I I D A W D M T I A Y G T D T I H P P D H V I G N Q I N 7484349  
 504 V G V V D A W D M T I A Y G T N N V H P P P D Q V I G N Q I N g1762

FIGURE 3C

WO 02/46411

18/18

PCT/US01/46767

3229449  
7484349  
91762

469 I L L N Y I C  
541 M F L N Y I C  
534 M F L N Y I C

FIGURE 3D

WO 02/46411

1/14

PCT/US01/46767

<110> INCYTE GENOMICS, INC.  
 LASEX, Amy W.  
 KRASNOW, Randi E.  
 BAUGHN, Mariah R.

<120> INTESTINAL PROTEINS

<130> PC-0028 PCT

<140> To Be Assigned  
 <141> Herewith

<150> 09/729,454  
 <151> 2000-12-04

<160> 32

<170> PERL Program

<210> 1  
 <211> 475  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3229449CD1

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

WO 02/46411

2/14

PCT/US01/46767

245  
 His Met Tyr Ser Lys Asn Lys Lys Val Ser Tyr Leu Ser Lys Gln 255  
 260  
 Glu Lys Ser Leu Phe Glu Arg Ser Asn Val Gly Val Glu Ile Met 270  
 275  
 Glu Lys Phe Asn Thr Ile Ser Val Ser Lys Cys Asn Thr Leu Lys 285  
 290  
 Ser Val Asp Leu His Glu Ser Gly Lys Leu Gln His Gln Leu Ala 300  
 305  
 Val Asp Leu Asp Arg Asn Ile Asn Ile Gln Trp Gln Lys Tyr Cys 315  
 320  
 Tyr Pro Leu Ile Gly Ser Met Thr Tyr Ser Val Lys Glu Met Glu 330  
 335  
 Tyr Leu Thr Arg Ala Ile Asp Arg Thr Gly Gly Glu Lys Asn Thr 345  
 350  
 Val Ile Val Ile Ser Leu Gly Gln His Phe Arg Pro Phe Pro Ile 360  
 365  
 Asp Val Phe Ile Arg Arg Ala Leu Asn Val His Lys Ala Ile Gln 375  
 380  
 His Leu Leu Leu Arg Ser Pro Asp Thr Met Val Ile Ile Lys Thr 390  
 395  
 Glu Asn Ile Arg Glu Met Tyr Asn Asp Ala Glu Arg Phe Ser Asp 405  
 410  
 Phe His Gly Tyr Ile Gln Tyr Leu Ile Ile Lys Asp Ile Phe Gln 420  
 425  
 Asp Leu Ser Val Ser Ile Ile Asp Ala Trp Asp Ile Thr Ile Ala 435  
 440  
 Tyr Gly Thr Asn Asn Val His Pro Pro Gln His Val Val Gly Asn 450  
 455  
 Gln Ile Asn Ile Leu Leu Asn Tyr Ile Cys 465  
 470  
 475

<210> 2  
 <211> 547  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7484349CD1

<400> 2  
 Met Ser Ser Asn Thr Met Leu Gln Lys Thr Leu Leu Ile Leu Ile  
 1 5 10 15  
 Ser Phe Ser Val Val Thr Trp Met Ile Phe Ile Ile Ser Gln Asn  
 20 25 30  
 Phe Thr Lys Leu Trp Ser Ala Leu Asn Leu Ser Ile Ser Val His  
 35 40 45  
 Tyr Trp Asn Asn Ser Ala Lys Ser Leu Phe Pro Lys Thr Ser Leu  
 50 55 60  
 Ile Pro Leu Lys Pro Leu Thr Glu Thr Glu Leu Arg Ile Lys Glu  
 65 70 75  
 Ile Ile Glu Lys Leu Asp Gln Gln Ile Pro Pro Arg Pro Phe Thr  
 80 85 90  
 His Val Asn Thr Thr Thr Ser Ala Thr His Ser Thr Ala Thr Ile  
 95 100 105  
 Leu Asn Pro Arg Asp Thr Tyr Cys Arg Gly Asp Gln Leu Asp Ile  
 110 115 120  
 Leu Leu Glu Val Arg Asp His Leu Gly Gln Arg Lys Gln Tyr Gly  
 125 130 135  
 Gly Asp Phe Leu Arg Ala Arg Met Ser Ser Pro Ala Leu Thr Ala  
 140 145 150  
 Gly Ala Ser Gly Lys Val Met Asp Phe Asn Asn Gly Thr Tyr Leu

WO 02/46411

3/14

PCT/US01/46767

155 160 165  
 Val Ser Phe Thr Leu Phe Trp Glu Gly Gln Val Ser Leu Ser Leu  
 170 175 180  
 Leu Leu Ile His Pro Ser Glu Gly Ala Ser Ala Leu Trp Arg Ala  
 185 190 195  
 Arg Asn Gln Gly Tyr Asp Lys Ile Ile Phe Lys Gly Lys Phe Val  
 200 205 210  
 Asn Gly Thr Ser His Val Phe Thr Glu Cys Gly Leu Thr Leu Asn  
 215 220 225  
 Ser Asn Ala Glu Leu Cys Glu Tyr Leu Asp Asp Arg Asp Gln Glu  
 230 235 240  
 Ala Phe Tyr Cys Met Lys Pro Gln His Met Pro Cys Glu Ala Leu  
 245 250 255  
 Thr Tyr Met Thr Thr Arg Asn Arg Glu Val Ser Tyr Leu Thr Asp  
 260 265 270  
 Lys Glu Asn Ser Leu Phe His Arg Ser Lys Val Gly Val Glu Met  
 275 280 285  
 Met Lys Asp Arg Lys His Ile Asp Val Thr Asn Cys Asn Lys Arg  
 290 295 300  
 Glu Lys Ile Glu Glu Thr Cys Gln Val Gly Met Lys Pro Pro Val  
 305 310 315  
 Pro Gly Gly Tyr Thr Leu Gln Gly Lys Trp Ile Thr Thr Phe Cys  
 320 325 330  
 Asn Gln Val Gln Leu Asp Thr Ile Lys Ile Asn Gly Cys Leu Lys  
 335 340 345  
 Gly Lys Leu Ile Tyr Leu Leu Gly Asp Ser Thr Leu Arg Gln Trp  
 350 355 360  
 Ile Tyr Tyr Phe Pro Lys Val Val Lys Thr Leu Lys Phe Phe Asp  
 365 370 375  
 Leu His Glu Thr Gly Ile Phe Lys Lys His Leu Leu Leu Asp Ala  
 380 385 390  
 Glu Arg His Thr Gln Ile Gln Trp Lys Lys His Ser Tyr Pro Phe  
 395 400 405  
 Val Thr Phe Gln Leu Tyr Ser Leu Ile Asp His Asp Tyr Ile Pro  
 410 415 420  
 Arg Glu Ile Asp Arg Leu Ser Gly Asp Lys Asn Thr Ala Ile Val  
 425 430 435  
 Ile Thr Phe Gly Gln His Phe Arg Pro Phe Pro Ile Asp Ile Phe  
 440 445 450  
 Ile Arg Arg Ala Ile Gly Val Gln Lys Ala Ile Glu Arg Leu Phe  
 455 460 465  
 Leu Arg Ser Pro Ala Thr Lys Val Ile Ile Lys Thr Glu Asn Ile  
 470 475 480  
 Arg Glu Met His Ile Glu Thr Glu Arg Phe Gly Asp Phe His Gly  
 485 490 495  
 Tyr Ile His Tyr Leu Ile Met Lys Asp Ile Phe Lys Asp Leu Asn  
 500 505 510  
 Val Gly Ile Ile Asp Ala Trp Asp Met Thr Ile Ala Tyr Gly Thr  
 515 520 525  
 Asp Thr Ile His Pro Pro Asp His Val Ile Gly Asn Gln Ile Asn  
 530 535 540  
 Met Phe Leu Asn Tyr Ile Cys  
 545

<210> 3  
 <211> 1616  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3229449CB1

WO 02/46411

4/14

PCT/US01/46767

```

<400> 3
atccatgcta aaggtaaaca aactgcaact tatatctgca atttatittg gtatagacaa 60
ggggtatgoc agtagcacac tggggcttcc agaagaatt ctcaacacct agctcgccag 120
agagtctatg tatgggattg acaaatctgt aaactaaagg atcctaataca tgaaaataag 180
tatgataaat tataagtcac tattggcact gttgtttata tttagcctcct ggatcatttt 240
tacagttttc cagaactcca caaagtttgg gtctgctcta aacttatcca tctcctccca 300
ttactggaac aactccacaa agtcttatt ccctaataca ccaactgata cattaaagcc 360
actaacagag actgaactca gaataaagga aatcatagag aaactagatc agcagatccc 420
accagacctc ttcaccacag tgaacaccac caccagcgcc acacatagca cagccaccat 480
ctcacaacct cgagatcagt actgcagggg agaccagctg cacatcctcg tggaggtgag 540
ggaccacttg ggacgcagga agcaatatgg oggggatttc ctgagggcca ggatgtcttc 600
cccagcgctg atggcaggtg cttcaggaaa ggtgactgac ttcacaacgc gcaactacct 660
ggtcagcttc actctgttct gggaggggcca ggtctctctg tctctgctgc tcatccacc 720
cagtgagggg gtgtcagctc tctggagtg aaggaaccac ggctatgaca ggggtactct 780
cactggccag ttgtcaatg gacttccca agtccactct gaatgtggcc tgatccataa 840
cacaaatgct gaatgtgccc agtacctgga cacagagagc caagaaggct tctactgtgt 900
ggggcctcaa cacatgccct gtctgactc cactcacatg tattctaaga caaagaaggt 960
ttcttatctt agcaaacacg asaagagcct otttgaaggg tcaaalgtgg gtgtagagat 1020
tatggaaaaa tcaatacaca ttagtgtctc caaatgcaac acactgaagt cagtggatct 1080
gcctgaatct gaaaaattgc aacaccagct tgctgtggat ttggatagga acatcaecat 1140
ccagtggcaa aatatattgt atcccttgat aggatcaatg acctattcag tcaaaagagat 1200
ggagtaccct acccggggcca tgcacagaac tggaggagaa aaaaatactg tcattgttat 1260
ttccctgggc cagcatttca gaccctttcc cattgatgtt tttatccgaa gggccctcaa 1320
tgtccacaaa gccattcagc atcttcttct gagaagccca gacactatgg ttatcatcaa 1380
aacagaaaaa atcagggaga tgtacaatga tgcagaaaga tttagtactc ttcattggtta 1440
cattcaatat ctcatcatala aggcattttt ccaggatctc agtgtgagta tcaattgatc 1500
ctgggatata acaattgcat atggcacaaa taatgtacac ccacctcaac atgtagtctg 1560
aaatcagatt aatatattat taaactatat ttgttaataa acaaaaaaaaa aaaaaa 1616

```

```

<210> 4
<211> 240
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 2771041H1

```

```

<400> 4
atccatgcta aaggtaaaca aactgcaact tatatctgca atttatittg gtatagacaa 60
ggggtatgoc agtagcacac tggggcttcc agaagaatt ctcaacacct agctcgccag 120
agagtctatg tatgggattg acaaatctgt aaactaaagg atcctaataca tgaaaataag 180
tatgataaat tataagtcac tattggcact gttgtttata tttagcctcct ggatcatttt 240

```

```

<210> 5
<211> 621
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 71851705V1

```

```

<400> 5
gccacattca gaagtggact tgggaagtgc cattgacaaa ctggccagtg aagatcacc 60
tgtcatagcc ttggttccct gcactccaga gagctgacac cccttccctg ggttgatga 120
gcagcagaga cagagagacc tggccctccc agaacagagt gaagctgacc aggtagggtc 180
cgtttgttga gtcagtcacc tttcctgaag cacctgacct cagcgtgggg gaagacatcc 240
tggccctcag gaaatccccg ccatattgct tccctgcgtc caagtgtgoc ctcaacctca 300
gcagagtgtg cagctgtgct cccctgcagt acgtatctcg aggttgagg atggtggcty 360
tgctatgtgt ggcctgggtg gtgggtttca cgtgggtgaa aggtctgggt ggggtctgct 420
gatctagttt ctctatgatt tcttttattc tgagttcagt ctctgttagt ggctttaatg 480
atatcagtggt tgttttaggg aataaggact ttgtggagtt gttccagtaa tggagggaga 540

```

WO 02/46411 5/14 PCT/US01/46767

tggataagtt tagagcagac caaacctttg tggagtctgt gaaaactgta aaaatgatcc 600  
aggaggctaa tataaacaac a 621

<210> 6  
<211> 545  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 70255975V1

<400> 6  
gccacattca gagtggactt ggggaagtgc attgacaaac tggccagtga agatcacctc 60  
gtcatagcct tggttccttg cactccagag agctgacacc ccttcaactg ggtggatgag 120  
cagcagagac agagagacct ggccctccca gaacagagtg aagctgacca ggtaggtgcc 180  
gtgtgtgaag tcagtccact ttcttgaagc acctgccatc agcgtggggg aagcatcctc 240  
ggccctcaag aaatccccgc catattgctt cctgcctccc aagtggctcc tcacctccag 300  
cagatgtgac agctggcttc cctgcagta cgtatctcga gggttgagga tgggtgctgt 360  
gctatgtgtg gcgctgtgtg tgggtttcac gtgggtgatc ggtctgggtg ggatctgctg 420  
atctagtttc tctatgattt cctttattct gagttcagtc tctgttagtg gctttaatga 480  
tatcagtggt gttttaggga ataaggactt tctggacgtc gttcagtaat ggaggggat 540  
ggata 545

<210> 7  
<211> 236  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 5596934H1

<220>  
<221> unsure  
<222> 228  
<223> a, t, c, g, or other

<400> 7  
ggaaccaagg ctattgacag ggtgatcttc actggccagt ttgtcaatgg cacttcccaa 60  
gtccactctg aatgtggcct gatcctaacc acaaatgctg aattgtgcca gtacctggac 120  
aacagagacc aagaaggctt ctactgtgtg aggcctcaac acatgccctg tgctgcactc 180  
actcacatgt attctaagaa caagaaggtt tottatctta gcaaacanga aaagag 236

<210> 8  
<211> 414  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 3229449F6

<220>  
<221> unsure  
<222> 47, 105, 248  
<223> a, t, c, g, or other

<400> 8  
caagtttggg atgacatcca caatccccag tgggcatgtc tggaganaca catggaatcc 60  
tgtctcctgt agtttggcta cagtcaaaat gaaggaatgc ctganaggaa aactcatata 120  
cctaagggga gattccaaga tccgcbagtg gatgaaatac ttcaagcca gtatcaacac 180  
actgaagtca gtggatctgc atgaatctgg aaaattgcaa caccagcttg ctgtggattt 240

WO 02/46411

6/14

PCT/US01/46767

```

ggataggac atcaacatcc agtggcaaaa acattggtat ccttgatag gatcaatgac 300
ctattcagtc aaagagatgg agtactctac cggggccat tgcacagact ggaggagaa 360
aaaaactctg tcaattgttat tccctgggg ccagcatttc agaccctttt ccca 414

```

```

<210> 9
<211> 394
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7128544H1

```

```

<400> 9
ggtaggttac tccatctctt tgacagaact ggaggagaaa aaaactctgt cattgtttat 60
tccctggggc agcatttccg accctttccc attgatgttt ttatccgaag ggccctcaat 120
gtccacaasg ccaatccaga tcttctctct agaagcccag acactatggt tatcatcaaa 180
acagaaaaaa tcaggggagat gtacaatgat gcagaaagat ttagtgactt tcatgtttac 240
attcaatctc tcatcataaa ggacattttc caggatctca gtgtgagtat cattgatgcc 300
tgggatataa caattgcata tggcacaat aatgtacacc cacctcaaca tgtagtccga 360
aatcagatta atatattatt aaactatatt tgatt 394

```

```

<210> 10
<211> 2248
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7484349CB1

```

```

<400> 10
gccagttaag aactccacgg tacttttcca gccaccaga taggagagat cattaaaaaa 60
gtgcattctg tgctacctga cacctattgg ggtcctggaa ggaggagca acaatcctga 120
gtgaaacctc gacaagaagt atccaatagg acattcgtca tgtcctcaaa tacaatgctt 180
caaaaaacgc tctgatcttt gatctctttt tcaagtagtaa cctggatgat ttttataatt 240
tctcagaact tcacaagact ttggtctgct ctaaacctat ccactctctg ccactactgg 300
aaccaactccg caaagtcctt attccctaaa acatcaactga taccattaaa gccactaaca 360
gagactgaec tcagaataaa ggaatcata gagaactag atcagcagat cccaccaga 420
cctttcacc c atgtgaacac caaccacagt gccacaaca gccacgccc catcctcaac 480
cctcgagata catactgcag gggagaccag ctggacatcc tactggaggt gagggccacc 540
ttggacaga ggaagcaata tgggtgggat ttcttgaggg ccaggatgtc ctcccagca 600
ctgacggcag gtgcttcagg aaaggtgatg gacttcaaca atggcacta cctggtcagc 660
ttcactctgt tctgggaggg caaggtctcc ctgtctctgc tgcctatcca ccccagtgaa 720
ggggctctgt ctctctggag gcaaggaac caaggtatg ataaaatta tttcaaggc 780
aaatttgta atggcaacct tcatgtcttc actgaatgtg gcctgacctt aaactcaaat 840
gtgtaacct gtgaatctct ggatgacaga gaccaagaag ccttctattg tatgaagcct 900
caacacatgc cctgtgaggc tctgacctac atgaccaacc ggaatagaga ggtatcttat 960
cttacagaca aggaanaacag ccttttccac aggtccaag tgggagttga aatgatgaag 1020
gatcgtaaac acattgatgt cactaattgt aacaagagag aaaaaataga agagacatgc 1080
caagtggaaa tgaagctcc tctcctcgtt ggttatactt tacaaggaaa atggataaca 1140
acattttgca accaggttca gttgacaca attaagata atgctgtttt gaaaggcaaa 1200
ctcatttacc tctgggaga ccttacaacta cgtcaagtga tctactactt ccccaagtt 1260
gtaaaacac tgaagttttt tgatcttcat gaaactggaa tctttaaga acatttgcct 1320
ctggatgcag aaagacacac tcagattcaa tggaaaaac atagctatcc ctctgctact 1380
ttccagctct acctctgat agatcatgat tatatccctc gggaaattga ccggctatca 1440
ggtgacaaaa acacagccat cgtcatcacc tttggcccag actttagacc atttccatt 1500
gacattttta tccgagggc catcgtgtgt caaaggcta ttgaaagact gttcctaaga 1560
agcccagcca ctaaatgtsat tattaagaca gaaaacatca gggagatgca catagagaca 1620
gagctggttg gagacttcca tggttatatt cactatctta tcatgagga tattttcaaa 1680
gacctcaacg tggcctcat tgatgctgg gacatgacca ttgcatatgg cactgacact 1740
atccaccac ctgatcatgt gattgaaat cagatgaaca tgttcttaa ctacatttc 1800
taagggataa tactatata aaactactag gaaccaatct ctgcacataa tcccacatgt 1860

```

WO 02/46411 7/14 PCT/US01/46767

atgtgaagt aagttttact ctttttagga actaaggaaa ataaatttaa aagaatcbgt 1920  
 ttggggagga aggtctatgta aggcacatga caactgataa gggatgcaaa accaaggaaa 1990  
 tcatctatga agaabgacta tcccatgctt ggttctgatg ctctgtttaa atattaaaa 2040  
 agtttttttaa aagccatggtt attaagctga ttgaaaata tctgtacaaa ttctgatgc 2100  
 ttctatttc caatatagat atttcttagc tctgtctatt gaaaaggcct aggcgcaatg 2160  
 ataacccatt agcaataatc actccgagca ccctaactgt gatgtctaag aacccttctt 2220  
 caataaaga aaagaggcat cettgaag 2248

<210> 11  
 <211> 661  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1333949F6

<400> 11  
 gccagttaag aactcacagg tacttttcca gcccccaga tagggagat cattaaaaca 60  
 gtgcattctg tctaacctga cacctattgg ggtcctggaa ggaggaagca acaatcctga 120  
 gtgaaacctc gacaagaagt atccaatagg acattcgtca tctcctcaaa tacaatgctt 180  
 caaaaaacgc tctgatcttt gatctctttt tcaagtatga cctggatgat tttataatt 240  
 tctcagaact tcacaagct ttggctctgt ctaaaactat ccactctgt ccattactgg 300  
 aacaactccg caaagtcctt attcctaaa acatcaactga taccattaaa gccactaaca 360  
 gagactgaac tcagaataaa ggaatcata gagsaactag atcagcagat ccccccaga 420  
 ctttccacc atgtgaacac caccaccagt gccacacaca gccagccac catcctcaac 480  
 cctcgaata catctgcaag gggagaccag ctggacatcc tactggggt gaggaacc 540  
 ttgggacaga ggaagcaata tgggtgggat ttctgaggg ccaggatgt cctcccaga 600  
 ctgcccagc gtgcttcagg aaagtgatg gactcaaca atggcaccta cctggctcgc 660  
 t 661

<210> 12  
 <211> 518  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7604658J1

<400> 12  
 ggtggtcatg taggtcagag cctcacagg catgtgtga ggcttcatac aatagaaggc 60  
 ttcttggtct ctgtcatoca gatattcaca gaggtaagca tttagttaa ggtcaggcc 120  
 acattcctgg aagcatgag aggtgccatt acaaatctt ccttgaaa taattttatc 180  
 atagccttgg ttccttgccc tccagagagc agacgcccc tcactggggt ggatgagcag 240  
 cagagacagg gagacctggc cctcccagaa cacagtgaag ctgaccaggt aggtgccatt 300  
 gttgaagctc atcaaccttc ctgaaacacc tgcctcagc gctggggagg acatcctggc 360  
 cctcaggaaa tccccccat attgcttct ctgtcccagg tggctcctca cctccagtag 420  
 gatgtccagc tggctctccc tgcagtatg atctcaggg ttgaggatgg tggctgtgct 480  
 gtgtgtgca ctggtggtgg tgttcacatg ggtgaaag 518

<210> 13  
 <211> 462  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 70106729V1

<400> 13  
 atagtgaata taaccatgga agtetccaaa cctctctgtc tctatgtgca tctcctgat 60  
 gttttctgtc ttaataatca ctttagtggc tgggttctt aggaacagtc tttcaatagc 120

WO 02/46411

8/14

PCT/US01/46767

```

cttttgaaca cggatggccc tgcgaataaa aatgtcaatg ggaatggtc taaagtgcgt 180
gcaaaaggtg atgacgatgg ctgtgttttt gtccactgat agccggfcaa ttcccggagg 240
gatataatca tgatctatca gtagatagag ctggaagtgt acgaagggat agctatgttt 300
ttccattgga atctgagtggt gtctttctgc atccagaagc aaatgtttct taaagattcc 360
agtttcatga agatcaaaaa aottcagtggt ttttacaact ttggggaagt agtagattca 420
ctgagctagt gtagagtctc ccaggaggtg aatgagtttg ct 462

```

<210> 14  
 <211> 531  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 70107804V1

```

<400> 14
ttttatttgc ttttaaaaaa cttttttaat attttaaacg agcatcagaa ccaggcatgg 60
tatagtcatt ctctcatgat gattctcttg gttttgcate ccttatcagt tgcctattgc 120
cttaccatagc ctctctcccc aaacagattc ttttaaatat atttccctta gttccataaaa 180
tgagtcaaac ttactttaca atacatgtgg gattatgtgc agagattggt tctatgtgat 240
tttgtatagt atttatocct tagcaaatgt agtttaagaa catgttcaac tgatttccaa 300
tcacatgatic aggtgggtgg atagtgtcag tgccatagc aatggtcatg tcccaggcat 360
caatgatgac caggttgagg tctttgaaaa tacccttcat gataagatag tgaatataac 420
catggaagtc tccaaacctc tctgtctcta tgtgcattcc cctgatgttt tctgtcttaa 480
taatacctct agtggctggg ctttttagga acagtcttcc aatagccttt t 531

```

<210> 15  
 <211> 276  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 5865314H1

<220>  
 <221> unsure  
 <222> 2, 15, 50  
 <223> a, t, c, g, or other

```

<400> 15
gnaaaaccaa gagantcatt catgaageat gactatacca tgccctggtt tgatgctcgt 60
ttaaataatt aaaaagttt tttaaaagcc atgttattaa gctgatttga aaatatctgt 120
acaaattcat gatgctttct atttcaata tagatatttc ctatgctctg ctattgaaaa 180
ggcctaggag caatgataac ccattagcaa taatcactcc gagcacccta actygtatgt 240
ctaagaacct ttcctcaata aaagaaaaga ggcatc 276

```

<210> 16  
 <211> 206  
 <212> DNA  
 <213> Rattus norvegicus

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 701244557H1

```

<400> 16
ccagccctga cggcaggcgc ttctggaaaa gtgacagact tcaacaatgg cgcctaccta 60
gtcagcttca ctctgctctg ggagggccag gtctccctgt ctatcctgct catgcacccc 120
agtgaagggg tgtcagctct ctggagagca aggaaccagg gttacgacag aatcatcttc 180
tcaggccatt ttgtcagtg cgtctc 206

```

WO 02/46411 9/14 PCT/US01/46767

<210> 17  
 <211> 291  
 <212> DNA  
 <213> Rattus norvegicus

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 700306567H1

<400> 17  
 cctggaagat attctttaag gcaagatact ggggtgaacc gtggaagtca ctaaactctc 60  
 ccattgctgtt attcaactcc ctgggtgttt ctggtttgag gaccaccagg gtgtccgggc 120  
 ttctctggag aagacgctga agagctctgt gaacactgag ggcctctgg ataaaaacat 180  
 caatgggaaa aggtctgaaa tgcctggcca gaaaaagac aatgactgtg tttttctctc 240  
 ctccgattct gtcaattatc cgtgcagtgt tctctatctc ttgacagag t 291

<210> 18  
 <211> 244  
 <212> DNA  
 <213> Rattus norvegicus

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 700141983H1

<400> 18  
 agcttccctc taatcgggtc attgggtgtac tctgtcaaaq agatagagaa cactgcacgg 60  
 ataattgaca gaalccggag agagaaaac acagtcattg tctttctctc gggccagcat 120  
 ttcagactct ttccattga tgtttttatc ggaaggccc tcaagtta cagagctctt 180  
 cagcgtcttc tccagagaag cccgcacacc ctggtgttcc tcaaacaga aaacaccagg 240  
 gagt 244

<210> 19  
 <211> 270  
 <212> DNA  
 <213> Rattus norvegicus

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 701725590H1

<400> 19  
 catggttcc ctctaactcg gtcattgggt tactctgtca aagagataga gaacactgca 60  
 cggataattg acagaatcgg aggagagaaa aacacagtca ttgtttttc tctgggccag 120  
 catttcagac cttttcccat tgatgttttt atccgaagg ccctcagtgt tcacagagct 180  
 ctccagctc tctccttag aagccgggac accctggtg tcctcaaac agaaaattat 240  
 agggagtga ataacgacat ggagaggtt 270

<210> 20  
 <211> 288  
 <212> DNA  
 <213> Rattus norvegicus

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 700363665H1

<400> 20  
 aacacagtca ttgtctttc tctgggccag catttcagac cttttccat tgatgtttt 60  
 atccgaacgg cctcagtgat tcacagagct ctccagctc ttctctgag aagcccggac 120  
 accctgggtg tctcaaac agaaaacacc atggagtga ataacgacat ggagaggtt 180  
 agtgattcca cgtttacacc cagtatcttg ccttaagaa tatcttcag gatctcctg 240  
 tgggtgtcat tgatgcctgg gatatgacag ttgcatatgg cacaaacg 288

WO 02/46411 10/14 PCT/US01/46767

<210> 21  
 <211> 275  
 <212> DNA  
 <213> Rattus norvegicus

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 701473585H1

<400> 21  
 gcagaccacaa cttgccgttg acttggatga gaaaatcaac atccagtggc agaaacatgg 60  
 cttccctcta atcgggtcat tgggttactc tgtcaaaagag atagagaaca ctgcacggat 120  
 aattgacaga atcggggag agaaaaaac agtcatlgtc ttttctctgg gccagcattt 180  
 cagacctttt cccattgatg ttttatocg ataggccctc agtgttcaca gagctcttca 240  
 ggtttcttc cagagaagcc cggacacctt ggtgg 275

<210> 22  
 <211> 257  
 <212> DNA  
 <213> Rattus norvegicus

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 700600759H1

<400> 22  
 gccaggtctc cctgtctatc ctgctcatgc accccaagta aggggtgtca gctctctgga 60  
 gggcagggaa ccaaggctat ggtagaatg ccttcaaaag gactttgtt aatggcacat 120  
 ccaaggtcac agctgaatgt ggcctgatcc tgaactcaag cagttagctc tgcaaatacc 180  
 tgtacctgtg tggcagggaa gtcttctact cgtgaaagcc tcaacacatg cctctgtggg 240  
 cctgacctca cgtgtgt 257

<210> 23  
 <211> 276  
 <212> DNA  
 <213> Rattus norvegicus

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 701460109H1

<400> 23  
 tatgggctgt gtttaagctg cctgcatcct tcaatcaatg ggacttgatc atgaaatcct 60  
 catgccctaa agtgcccttc aatccatcag tttccaacac agagacagaa ctgagaatca 120  
 aggagatcct agagaaacta acaaacaga tccctccag acccttcgcc caccitcaaca 180  
 acaccacaag tctacacac agcatagcca caatcctcaa ccctcaagat acatactgtg 240  
 taggggacca gctggacatc ctggtagagg ctagag 276

<210> 24  
 <211> 250  
 <212> DNA  
 <213> Rattus norvegicus

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 701420417H1

<400> 24  
 aacaaggaca tttcttatct tagccagcag gaaaggagcc tctttgaaag gtcaaatata 60  
 gctgtggaga ttatgggaaa atccaacgtg attagtgtct ccaaatgcaa caaagccgtc 120  
 ccggtgaaag agaatgcaa gtttgggatg gcatctgcaa tccctactgg gcatgtctgg 180  
 aaaaacagct ggaatccggc ctccctgcagt ctggctccaa tcaaatgaa agactgtctg 240  
 agaggaaaac 250

WO 02/46411 11/14 PCT/US01/46767

<210> 25  
 <211> 248  
 <212> DNA  
 <213> Rattus norvegicus

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 701634496H1

<400> 25  
 cggacaccct ggtggtcctc aaaacagaaa acaccagga gttgaataac gacatggaga 60  
 gggttagtga ctccacggt tacaccaggt atcttgccct aaagaatata tcccaggatc 120  
 tccggtgagg tgcattgat gcctgggata tgacagttgc atatggcaca aacgatgtcc 180  
 atccaccaga ggaggtagtt agaagtgaaa ttaatatatt cttaaaacta atttgctagc 240  
 aaacacat 248

<210> 26  
 <211> 329  
 <212> DNA  
 <213> Rattus norvegicus

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 701601584H1

<400> 26  
 ttaaagaata tcttcagga tctccgtgtg ggtgctattg atgcctgga tatgacagtt 60  
 gcataatgga caaacgatgt ccatccacca gagaggtag ttagaagtga aabtaabata 120  
 ttcttaaaact atatttgcta gcaaacacat aactttgaaa gtcgctcgtt gaacttaaaa 180  
 gagacagtga gtccctacagc cgtgcccaagt gccgagatat cccagttaat ccaaggacat 240  
 aatctgtatt atggtccatg tggctccatcc agttcagcct aataaggcat tcctacgcca 300  
 gcctgctgct caaaattgaa tatgaaaag 329

<210> 27  
 <211> 144  
 <212> DNA  
 <213> Rattus norvegicus

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 701940254H1

<400> 27  
 agcctccttg aaaggtcaaa tatagctgtg gagattatgg gaaaatccaa cgtgattagt 60  
 gtcctcaaat gcaacaggtt ctttgaaaaa gatggaaggc acttaataaa cacagatgaa 120  
 ctggtgtttt agaagacccc atct 144

<210> 28  
 <211> 262  
 <212> DNA  
 <213> Rattus norvegicus

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 701463630H1

<400> 28  
 gggaaaaggt ctgaaatgct ggcccagaga aaagacaatg actgtgtttt tctctcctcc 60  
 gattctgtca attatccgtg cagtgttctc tatctctttg acagagtaca ccaatgaccc 120  
 gattagagg aagccatggt tctgccaactg gatgttgalt ttctcatcca agtccaagcg 180  
 aagttggtgc tgcagccttc cagtctcgtg gaggtccacc ggcttcagcg tgttgatttt 240  
 gcttttgaag tactccatcc ac 262

WO 02/46411 12/14 PCT/US01/46767

<210> 29  
 <211> 277  
 <212> DNA  
 <213> Rattus norvegicus

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 701623610H1

<400> 29  
 ctcaacagga gatcctggaa gatattcttt aaggcaagat actgggtgta tccgtggaag 60  
 tccctgatgc ctgggatatg acagtggcat atggcacaaa cgatgtccat ccaccagagg 120  
 aggtagttag aagtgaaatt aatatattct taaactatat ttgctagcaa acacataact 180  
 ttgaaagtgc ctggttgaaac ttaaaagaga cagtggatcc tacagccgtg ccaagtgcg 240  
 agatattcca gttaatacaa ggacataatc tgratta 277

<210> 30  
 <211> 1005  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: GNM.g9965027\_000007\_006

<400> 30  
 atgtcttccc cagcgtgat ggcaggtgct tcaagaaagg tgactgactt caacaacggc 60  
 acctacctgg tcaagctcac tctgttctgg gaggccagg tctctctgtc tctgtctctc 120  
 atccaccoca gtgaaggggt gtcagctctc tggagtcaa ggaaccaagg ctatgcagg 180  
 gtgatcttca ctggccagtt tgcacatggc acttcccaag tccactctga atgtggcctg 240  
 atctcaaaa caaatgctga attgtgccag taccctggaca acagagacca agaagggctc 300  
 tactgtgta ggcctcaaca catgccctgt gctgcactca ctcaatgta ttctaagaac 360  
 aagaaagtgt cttatcttag caaacaagaa aagagcctct ttgaaaggtc aaatgtgggt 420  
 gtgagatta tggaaaaatt caatacaatt agtgtctcca aatgcaacac actgaaagta 480  
 gtggatctgc atgaatctgg aaaaatgcaa caccagcttg ctgtggattt ggataggaac 540  
 atcaacatcc agtggcaaaa atattgttat ccttggatag gatcaatgac clattcagtc 600  
 aaagagatgg agtacctcac cggggccatt gacagaactg gaggagaaaa aaatactgtc 660  
 atgttattt ccttgggcca gcaatcaga cctttccca ttgatgtttt tatccgaagg 720  
 gacctcaatg tccacaagc catctcgcat ctctcttctg gaagccaga gactatgggt 780  
 atcatcaaaa cagaacaact cagggagatg tacaatgatg cagaagatt tagtgacttt 840  
 catgttaca ttcaatatct catcataaag gacattttcc aggtctcag tgtgagtatc 900  
 attgatgctt gggatatac aattgcaat ggcaaaaata atgtacacc acctcaact 960  
 gtatcggaac atcagattaa tatattatta aactatattt gttaa 1005

<210> 31  
 <211> 1545  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: GNM.g9795680\_006.edit

<400> 31  
 ctttggcttg ctctaaactt atccatctct gtccattact ggaacaactc cgcgaagtcc 60  
 ttattcccta aaacatcaat gataccatta aagccaacta cagagactga actcagaata 120  
 aaggaatatc tagaagaact agatcagcag atcccaccca gaccttccac ccatgtgaac 180  
 accaccacca gtgccacaca cagcacagcc acctactcca accttcgaga tacatactgc 240  
 aggggagacc agctggacat cctactggag gtgagggacc acttgggaca gaggaaagca 300  
 tatgtgggg atttctcag ggcaggatg tcttcccag cactgacggc aggtgcttca 360  
 gaaaaggtag tggactcaa caatggcacc tacttggta gcttcaactt gttctggag 420  
 gcccaggtct cctgtctct gctgctcacc caccacagt aaggggctc ggtctctctg 480  
 agggcaagga accaaggcta tgataaaatt attttcaag gcaaatgtt taatggcacc 540

WO 02/46411

13/14

PCT/US01/46767

```

ttctatgtct tcaactgaatg tggcctgacc ctaaaactcaa atgctgaact ctgtgaatat 600
ctggatgaca gagaccaaga agccttctat tgtatgaagc ctcaacacat gccctgtggg 650
gtctctgacct acatgacac cgggaataga gaggatctct atcttacaga caaggaaaac 720
agccttttcc acaggtccaa agtgggagtt gaatgatga aggatcgtaa acacattgat 780
gtactaatl gtaacaagag agaaaaata gaagagacat gcccaagttg aatgaagcct 840
cctgtccctg gtggttatac tttacaagga aatggataa caacatttg caaccaggtt 900
cagttagaca caattaagat aatggctgt ttgaaaggca aactcatta cctcctggga 960
gactctaac tacgtcagtg gatctactc ttcccacaag ttgtaaaaac actgaagttt 1020
tttgatcttc atgaaactgg aatctttaag aaacatttgc ttctggatgc agaaagcac 1080
actcagatcc aatgaaaaaa acatagctat cccttcgtca ctttccagct ctactctctg 1140
atagatcatg attatatccc tcgggaaatt gaccggctat caggtgacaa aaacacagcc 1200
atogtcatca cctttggcca gcactttaga ccatttccca ttgacatttt tattgcagg 1260
gccatcggtg ttcaaaaaggc tattgaaaga ctgttccctaa gaagcccagc cactaaagtg 1320
attattaaga cagaaaacat cagggagatg cacatagaga cagagaggtt tggagacttc 1380
catggttata tccaatctct tatcatgaag catattttca aagacctcaa cgtgggcctc 1440
atgtatgctt gggacatgac cattgcaat ggcactgaca ctatccacc accatgatcat 1500
gtgattggaa atcagattaa catgttctta aactacattt gctaa 1545

```

```

<210> 32
<211> 540
<212> PRT
<213> Oryctolagus cuniculus

```

```

<220>
<221> misc_feature
<223> ineyte ID No: g1762

```

```

<400> 32
Met Leu His Lys Tyr Leu Lys Leu Ile Cys Leu Leu Ala Ala Ile
1 5 10 15
Cys Val Leu Cys Ile Ile Ser Gln Asn Ser Thr Lys Ile Trp Gly
20 25 30
Ala Leu Lys Leu Pro Asn Ser His Tyr Tyr Ser Asn Thr Ser Met
35 40 45
Ile Ser Ser Ile Pro Lys Met Ser Val Ser Pro Val Lys Ser Leu
50 55 60
Thr Glu Thr Glu Leu Arg Val Lys Glu Ile Leu Glu Lys Leu Asp
65 70 75
Arg Leu Ile Pro Pro Arg Pro Phe Thr His Val Asn Thr Thr Thr
80 85 90
Ser Ala Thr His Ser Thr Ala Thr Ile Leu Asn Pro Gln Asp Lys
95 100 105
Tyr Cys Val Gly Asp Gln Leu Asp Ile Leu Leu Glu Val Arg Asp
110 115 120
Tyr Leu Gly His Gln Lys Glu Tyr Gly Gly Asp Phe Leu Arg Ala
125 130 135
Arg Met Phe Ser Pro Ala Leu Lys Ala Gly Ala Ser Gly Lys Val
140 145 150
Thr Asp Phe Asn Asn Gly Thr Tyr Leu Val Ser Phe Thr Leu Phe
155 160 165
Trp Glu Gly Gln Val Ser Leu Ser Val Leu Leu Ile His Pro Ser
170 175 180
Glu Gly Ala Ser Ala Leu Trp Arg Ala Arg Asn Gln Gly Tyr Asp
185 190 195
Arg Ile Ile Phe Lys Gly Gln Phe Val Asn Gly Thr Ser His Val
200 205 210
Phe Thr Glu Cys Ser Leu Thr Leu Asn Ser Asn Thr Glu Glu Cys
215 220 225
Lys Tyr Leu Asn Gly Arg Asp Gln Asp Val Phe Tyr Cys Met Lys
230 235 240
Pro Gln His Met Pro Cys Glu Ala Leu Thr His Val Thr Ser Arg
245 250 255
Asn Arg Asp Ile Ser Tyr Leu Thr Ser Lys Glu Lys Asn Leu Phe

```

WO 02/46411

14/14

PCT/US01/46767

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

【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 13 June 2002 (13.06.2002)

PCT

(10) International Publication Number WO 2002/046411 A3

(51) International Patent Classification: C12N 15/12, C07K 14/47, 16/18, C12Q 1/68, G01N 33/68, A01K 67/027, A61K 39/395

(74) Agents: STREETER, David, G. et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(21) International Application Number: PCT/US2001/046767

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(22) International Filing Date: 3 December 2001 (03.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data: 09/729,454 4 December 2000 (04.12.2000) 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, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

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

Published: with international search report

(72) Inventors; and (75) Inventors/Applicants (for US only): LASEK, Amy, W. [US/US]; 237 41st Street #5, Oakland, CA 94611 (US); KRASNOW, Randi [US/US]; 817 Santa Fe Avenue, Stanford, CA 94305 (US); BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US).

(88) Date of publication of the international search report: 12 February 2004

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.

(54) Title: INTESTINAL PROTEINS

5' A TCC ATG CTA AAG GTA AAC AAA CTG CAA CTT ATA TCT GCA ATT TAT TTT GGT
63 63 72 81 90 99 108
ATA GAC ARG AGG TAT GCC AGT AGC ACA CTG GTG GCT TCA GAA GAA ATT CTC AAC
117 117 126 135 144 153 162
ACC TAG CTC GCC AGA GAG TCT ATG TAT GGG ATT GAA CAA TCT GTA AAC TAA AGG
171 171 180 189 198 207 216
ATC CTA ATC ATG AAA ATA AAT ATG ATA AAT TAT AAG TCA CTA TTG GCA CTG TTG
M K I S M I N Y K S L L A L L
225 225 234 243 252 261 270
TTT ATA TTA GCC TCC TGG ATC ATT TTT ACA GTT TTC CAG AAC TCC ACA AAG GTT
F I L A S W I I F T V F Q N S T K V
279 279 288 297 306 315 324
TGG TCT GCT CTA AAC TTA TCC ATC TCC CTC CAT TAC TGG AAC AAC TCC ACA AAG
W S A L N L S I S L H Y W N N S T K

(57) Abstract: The invention provides cDNAs which encode mammalian intestinal proteins. It also provides for the use of the cDNAs, fragments, complements, and variants thereof and of the encoded protein, portions thereof and antibodies thereto for diagnosis and treatment of colon disorders, particularly colon cancer, Crohn's disease, and ulcerative colitis. The invention additionally provides expression vectors and host cells for the production of the protein and a transgenic model system.

WO 2002/046411 A3

## 【国際調査報告】

INTERNATIONAL SEARCH REPORT		International Application No. PCT/US 01/46767
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/12 C07K14/47 C07K16/18 C12Q1/68 G01N33/68 A01K67/027 A61K39/395		
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)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BOLL WERNER ET AL: "Messenger RNAs expressed in intestine of adult but not baby rabbits: Isolation of cognate cDNAs and characterization of a novel brush border protein with esterase and phospholipase activity" JOURNAL OF BIOLOGICAL CHEMISTRY, AMERICAN SOCIETY OF BIOLOGICAL CHEMISTS, BALTIMORE, MD, US, vol. 268, no. 17, 15 June 1993 (1993-06-15), pages 12901-12911, XP002200728 ISSN: 0021-9258 page 12904, column 1, paragraph 3 -& DATABASE EMBL 'Online! 19 November 1992 (1992-11-19) "O. cuniculus mRNA for protein of unknown function" Database accession no. Z12840 -/-	1,4-15
X		1,4-15
<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:		
*A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art *Z* document member of the same patent family		
Date of the actual completion of the international search	Date of mailing of the international search report	
20 November 2002	06.03.03	
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patenlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer  Weikl, M	

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

INTERNATIONAL SEARCH REPORT		International Application No PCT/US 01/46767
C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>XP002220545 77.9% identity with SEQ ID No 3 in 906 bp overlap and 83.6% identity with SEQ ID No 10 in 1684 bp overlap the whole document -&amp; DATABASE SMALL 'Online! 1 February 1994 (1994-02-01) "Brush border 61.9 kDa protein precursor" Database accession no. Q05004 XP002220546 cited in the application 57.8% identity with SEQ ID No 1 in 536 aa overlap and 78.8% identity with SEQ ID No 2 over its entire length (540 aa) the whole document</p>	1,4-15
X	<p>WO 00 12702 A (BURGESS CHRISTOPHER C ;CARROLL EDDIE III (US); CATINO THEODORE J ( ) 9 March 2000 (2000-03-09) the whole document</p>	1-31
A	<p>SCHWEINFEST C W ET AL: "Subtraction hybridization cDNA libraries from colon carcinoma and hepatic cancer" GENE ANALYSIS TECHNIQUES, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 7, 1990, pages 64-70, XP002089887 ISSN: 0735-0651 the whole document</p>	1-31
A	<p>VIDER B ET AL: "Human colorectal carcinogenesis is associated with deregulation of homeobox gene expression" BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, ACADEMIC PRESS INC. ORLANDO, FL, US, vol. 232, no. 3, March 1997 (1997-03), pages 742-748, XP002104685 ISSN: 0006-291X the whole document</p>	1-31

From PCT/BAZ1R (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT	
International application No. PCT/US 01/46767	
<b>Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)</b>	
This International Search Report has not been established in respect of certain claims under Article 17(2)(e) for the following reasons:	
1.	<input checked="" type="checkbox"/> Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely: see FURTHER INFORMATION sheet PCT/ISA/210
2.	<input checked="" type="checkbox"/> Claims Nos.: 1c, e; 16c, e because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically: see FURTHER INFORMATION sheet PCT/ISA/210
3.	<input type="checkbox"/> Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
<b>Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)</b>	
This International Searching Authority found multiple inventions in this International application, as follows:	
1.	<input type="checkbox"/> As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2.	<input type="checkbox"/> As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3.	<input type="checkbox"/> As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4.	<input checked="" type="checkbox"/> No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1, 2, 3a, b, d(partially), 4-31
<b>Remark on Protest</b>	
	<input type="checkbox"/> The additional search fees were accompanied by the applicant's protest.
	<input type="checkbox"/> No protest accompanied the payment of additional search fees.

## INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 01 /46767

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claim 31 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.  
Although claims 13 and 28 are directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Continuation of Box I.2

Claims Nos.: 1c,e; 16c,e

Present claims 1c and 16c relate to a protein fragment defined by reference to a desirable characteristic or property, namely antigenic activity.

The claims cover all protein fragments having this characteristic or property, whereas the application provides support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT for only a very limited number of such fragments. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT). An attempt is made to define the protein fragment by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search over the whole of the claimed scope impossible. Consequently, the search has been carried out for those parts of the claims which appear to be clear, supported and disclosed, namely those parts relating to the protein fragments T66-H90, L120-R141, D234-R245 and I351-N359 of IP-1 and T104-R140 and R194 -K206 of IP-2.

Present claims 1e and 16e also relate to a protein fragment defined by reference to a desirable characteristic or property, namely biological activity.

The claims cover all protein fragments having this characteristic or property, whereas the application provides no support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT for such fragments. This is of course also due to the fact that the present application fails to assign a physiological/biological function to the respective proteins. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT). An attempt is made to define the protein fragments by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search over the whole of the claimed scope impossible. Consequently, no search has been carried out for these claims.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an

## INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 01 A6767

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

## INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 01 A6767

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1, 2, 3a, b, d(partially), 4-31  
Claims relating to protein (SEQ ID No 1 and 2) and DNA sequences (SEQ ID No 3 and 10) of intestinal proteins IP-1 and IP-2; also claims relating to fragments of these DNA sequences (SEQ ID Nos 4-9 and 11-15)
2. Claim : 3c and d (insofar as they relate to SEQ ID No 16)  
claims relating to a variant of SEQ ID Nos 3 and 10 where alleged function has not been demonstrated
3. Claim : 3c and d (insofar as they relate to SEQ ID No 17)  
claims relating to a variant of SEQ ID Nos 3 and 10 where alleged function has not been demonstrated
4. Claim : 3c and d (insofar as they relate to SEQ ID No 18)  
claims relating to a variant of SEQ ID Nos 3 and 10 where alleged function has not been demonstrated
5. Claim : 3c and d (insofar as they relate to SEQ ID No 19)  
claims relating to a variant of SEQ ID Nos 3 and 10 where alleged function has not been demonstrated
6. Claim : 3c and d (insofar as they relate to SEQ ID No 20)  
claims relating to a variant of SEQ ID Nos 3 and 10 where alleged function has not been demonstrated
7. Claim : 3c and d (insofar as they relate to SEQ ID No 21)  
claims relating to a variant of SEQ ID Nos 3 and 10 where alleged function has not been demonstrated
8. Claim : 3c and d (insofar as they relate to SEQ ID No 22)  
claims relating to a variant of SEQ ID Nos 3 and 10 where alleged function has not been demonstrated
9. Claim : 3c and d (insofar as they relate to SEQ ID No 23)  
claims relating to a variant of SEQ ID Nos 3 and 10 where alleged function has not been demonstrated

## INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 01 A6767

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

10. Claim : 3c and d (insofar as they relate to SEQ ID No 24)  
claims relating to a variant of SEQ ID Nos 3 and 10 where  
alleged function has not been demonstrated
11. Claim : 3c and d (insofar as they relate to SEQ ID No 25)  
claims relating to a variant of SEQ ID Nos 3 and 10 where  
alleged function has not been demonstrated
12. Claim : 3c and d (insofar as they relate to SEQ ID No 26)  
claims relating to a variant of SEQ ID Nos 3 and 10 where  
alleged function has not been demonstrated
13. Claim : 3c and d (insofar as they relate to SEQ ID No 27)  
claims relating to a variant of SEQ ID Nos 3 and 10 where  
alleged function has not been demonstrated
14. Claim : 3c and d (insofar as they relate to SEQ ID No 28)  
claims relating to a variant of SEQ ID Nos 3 and 10 where  
alleged function has not been demonstrated
15. Claim : 3c and d (insofar as they relate to SEQ ID No 29)  
claims relating to a variant of SEQ ID Nos 3 and 10 where  
alleged function has not been demonstrated

**INTERNATIONAL SEARCH REPORT**  
Information on patent family members

International Application No  
PCT/US 01/46767

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0012702 A	09-03-2000	AU 5691699 A	21-03-2000
		EP 1108021 A	20-06-2001
		JP 2002523088 T	30-07-2002
		US 6262334 B	17-07-2001
		US 2002034758 A	21-03-2002

## フロントページの続き

(51) Int.Cl. <sup>7</sup>	F I	テーマコード(参考)
C 0 7 K 16/18	C 0 7 K 14/47	4 C 0 8 5
C 1 2 N 1/15	C 0 7 K 16/18	4 H 0 4 5
C 1 2 N 1/19	C 1 2 N 1/15	
C 1 2 N 1/21	C 1 2 N 1/19	
C 1 2 N 5/10	C 1 2 N 1/21	
C 1 2 P 21/02	C 1 2 P 21/02	C
C 1 2 P 21/08	C 1 2 P 21/08	
C 1 2 Q 1/68	C 1 2 Q 1/68	A
G 0 1 N 33/15	G 0 1 N 33/15	Z
G 0 1 N 33/50	G 0 1 N 33/50	Z
G 0 1 N 33/53	G 0 1 N 33/53	D
G 0 1 N 33/566	G 0 1 N 33/53	M
	G 0 1 N 33/566	
	C 1 2 N 5/00	A

(81) 指定国 AP(GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP(AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, 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 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 AA34 AA35 AA40 BA11 BB50 DA13 DA36 FB02 FB03  
 4B024 AA01 AA11 AA12 BA44 BA80 CA04 CA07 DA12 EA04 GA11  
 HA12  
 4B063 QA18 QA19 QQ42 QQ52 QR55 QR62 QR82 QS25 QS34  
 4B064 AG01 CA06 CA19 CC24 DA01 DA05 DA13 DA14  
 4B065 AA72X AA93X AB01 AC14 BA02 CA24 CA44 CA46  
 4C085 AA13 AA14 BB01 BB11 CC23 CC32 EE01  
 4H045 AA10 AA11 AA20 AA30 BA10 CA40 DA76 DA86 EA20 EA22  
 EA28 EA50 EA51 FA71 FA74 GA26

专利名称(译)	肠道蛋白质		
公开(公告)号	<a href="#">JP2005505232A</a>	公开(公告)日	2005-02-24
申请号	JP2002548128	申请日	2001-12-03
[标]申请(专利权)人(译)	洞察Genomics公司		
申请(专利权)人(译)	洞察基因组公司		
[标]发明人	ラセクアミーダブリュ クラスノーランディ ボーグンマライアール		
发明人	ラセク、アミーダブリュ クラスノー、ランディ ボーグン、マライアール		
IPC分类号	G01N33/50 A61K39/395 A61P35/00 C07K1/22 C07K14/47 C07K16/18 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N15/09 C12N15/12 C12P21/02 C12P21/08 C12Q1/68 G01N33/15 G01N33/53 G01N33/566		
CPC分类号	C07K14/47 A61K2039/505		
FI分类号	C12N15/00.ZNA.A A61K39/395.E A61K39/395.T A61P35/00 C07K1/22 C07K14/47 C07K16/18 C12N1/15 C12N1/19 C12N1/21 C12P21/02.C C12P21/08 C12Q1/68.A G01N33/15.Z G01N33/50.Z G01N33/53.D G01N33/53.M G01N33/566 C12N5/00.A		
F-TERM分类号	2G045/AA34 2G045/AA35 2G045/AA40 2G045/BA11 2G045/BB50 2G045/DA13 2G045/DA36 2G045/FB02 2G045/FB03 4B024/AA01 4B024/AA11 4B024/AA12 4B024/BA44 4B024/BA80 4B024/CA04 4B024/CA07 4B024/DA12 4B024/EA04 4B024/GA11 4B024/HA12 4B063/QA18 4B063/QA19 4B063/QQ42 4B063/QQ52 4B063/QR55 4B063/QR62 4B063/QR82 4B063/QS25 4B063/QS34 4B064/AG01 4B064/CA06 4B064/CA19 4B064/CC24 4B064/DA01 4B064/DA05 4B064/DA13 4B064/DA14 4B065/AA72X 4B065/AA93X 4B065/AB01 4B065/AC14 4B065/BA02 4B065/CA24 4B065/CA44 4B065/CA46 4C085/AA13 4C085/AA14 4C085/BB01 4C085/BB11 4C085/CC23 4C085/CC32 4C085/EE01 4H045/AA10 4H045/AA11 4H045/AA20 4H045/AA30 4H045/BA10 4H045/CA40 4H045/DA76 4H045/DA86 4H045/EA20 4H045/EA22 4H045/EA28 4H045/EA50 4H045/EA51 4H045/FA71 4H045/FA74 4H045/GA26		
优先权	09/729454 2000-12-04 US		
外部链接	<a href="#">Espacenet</a>		

#### 摘要(译)

本发明提供了编码哺乳动物肠蛋白的cDNA。本发明还提供了其cDNA利用的片段，互补物和变体，并且可以用于编码蛋白质的利用，特别是用于结肠癌，库仑病和溃疡性结肠炎。提供自身的一部分和抗体，用于诊断和治疗结肠疾病。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。

