

(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2005-505267

(P2005-505267A)

(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)
<b>C 1 2 N 15/09</b>	C 1 2 N 15/00 Z N A A	2 G O 4 5
<b>A 6 1 K 31/7088</b>	A 6 1 K 31/7088	4 B O 2 4
<b>A 6 1 K 31/7105</b>	A 6 1 K 31/7105	4 B O 2 9
<b>A 6 1 K 31/711</b>	A 6 1 K 31/711	4 B O 6 3
<b>A 6 1 K 31/7115</b>	A 6 1 K 31/7115	4 B O 6 4
審査請求 未請求 予備審査請求 未請求 (全 141 頁) 最終頁に続く		

(21) 出願番号	特願2003-523619 (P2003-523619)	(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94304・パロアルト・ポータードライブ 3160
(86) (22) 出願日	平成14年8月22日 (2002.8.22)	(74) 代理人	100089266 弁理士 大島 陽一
(85) 翻訳文提出日	平成16年2月23日 (2004.2.23)	(72) 発明者	トリボレー、キャサリーン・エム アメリカ合衆国カリフォルニア州94107・サンフランシスコ・#5・テネシーストリート 1121
(86) 国際出願番号	PCT/US2002/027144		
(87) 国際公開番号	W02003/018768		
(87) 国際公開日	平成15年3月6日 (2003.3.6)		
(31) 優先権主張番号	60/314, 914		
(32) 優先日	平成13年8月24日 (2001.8.24)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 癌において差次的に発現する膜貫通タンパク質

## (57) 【要約】

本発明は、大腸癌または胃癌において差次的に発現する膜貫通タンパク質であるTMDCを提供する。また、このタンパク質、該タンパク質をコードするcDNA、および該タンパク質に特異的に結合する抗体類を種々の方法に用いて大腸癌または胃癌の進行または治療を診断、病期分類、または監視する方法をも提供する。

## 【特許請求の範囲】

## 【請求項 1】

以下から選択された或るタンパク質をコードする或る単離されたcDNA。

(a) SEQ ID NO:1 (配列番号1) のアミノ酸配列

(b) SEQ ID NO:1の抗原エピトープ、および

(c) SEQ ID NO:1のアミノ酸配列に対して少なくとも90%の配列同一性を有するアミノ酸配列

## 【請求項 2】

以下から選択された或るポリヌクレオチドを有する或る単離されたcDNA。

(a) SEQ ID NO:2の核酸配列、またはSEQ ID NO:2の相補配列

(b) SEQ ID NO:2の断片、またはその相補配列、および

(c) SEQ ID NO:2の核酸配列に対して少なくとも90%の配列同一性を有する或るポリヌクレオチド、またはその相補配列

10

## 【請求項 3】

SEQ ID NO:2の核酸配列を持つポリヌクレオチドを有する或る単離されたcDNA、または該cDNAの或る相補配列。

## 【請求項 4】

請求項2のcDNAを有する或るプローブ。

## 【請求項 5】

請求項2のcDNAで形質転換した或る細胞。

20

## 【請求項 6】

請求項2に記載のcDNAと、或る標識成分とを有する組成物。

## 【請求項 7】

請求項2のcDNAを有する或るアレイエレメント。

## 【請求項 8】

請求項2のcDNAがその上に固定された或る基板。

## 【請求項 9】

請求項2に記載のcDNAを有する或るベクター。

## 【請求項 10】

請求項9に記載のベクターを有する或る宿主細胞。

30

## 【請求項 11】

或るcDNAを用いてタンパク質を産生する方法であり、以下の過程を有する方法。

(a) 請求項10の宿主細胞の培養を、タンパク質発現に適した条件で行う過程、および

(b) 該タンパク質を該宿主細胞培地から回収する過程

## 【請求項 12】

請求項3に記載のcDNAと、或る標識成分とを有する或る組成物。

## 【請求項 13】

或るcDNAを用いてサンプル中の或る核酸の発現を検出する方法であり、

(a) 少なくとも1つのハイブリダイゼーション複合体が形成される条件下で請求項6の組成物と該サンプルの核酸群とをハイブリダイズする過程と、

(b) ハイブリダイゼーション複合体形成を検出する過程とを含み、複合体形成が、前記サンプル中の該核酸の発現を標示することを特徴とする方法。

40

## 【請求項 14】

ハイブリダイゼーション前にサンプル中の該核酸を増幅する過程をも有する請求項13に記載の方法。

## 【請求項 15】

組成物が或る基板に接着されることを特徴とする請求項13に記載の方法。

## 【請求項 16】

請求項13の方法であって、該サンプルが大腸または胃から得られる方法。

## 【請求項 17】

50

請求項13の方法であって、複合体形成が標準と比較され、大腸癌または胃癌の診断となる方法。

【請求項18】

或るcDNAを用いて複数の分子または化合物をスクリーニングする方法であり、以下の過程を有する方法。

(a) 特異結合を許容する条件下で、請求項2に記載のcDNAを複数の分子または化合物と混合する過程と、

(b) 特異結合を検出し、それにより該cDNAに特異結合する分子または化合物を同定する過程

【請求項19】

請求項18の方法であって、該分子群ないし化合物群が、アンチセンス分子、人工染色体作成物、分岐核酸、DNA分子、エンハンサー、ペプチド核酸、ペプチド、タンパク質、リプレッサ、RNA分子、および転写因子から選択される方法。

【請求項20】

或るcDNAを用いて或る分子または化合物の効力を算定する方法であって、以下の過程を含む方法。

(a) 核酸群を有する或るサンプルを該分子または化合物で処理する過程

(b) ハイブリダイゼーション複合体の形成に適した諸条件下で請求項2のcDNAを有する該サンプルの核酸群をハイブリダイズする過程

(c) 複合体形成の量を判定する過程

(d) 前記処理されたサンプル中の複合体形成量を、或る処理されないサンプル中の複合体形成量と比較する過程であって、複合体形成における差が、該分子または化合物の効力を標示することを特徴とする過程

【請求項21】

或るcDNAを用いて或る分子または化合物の毒性を算定する方法であって、以下の過程を含む方法。

(a) 核酸群を有する或るサンプルを該分子または化合物で処理する過程

(b) ハイブリダイゼーション複合体の形成に適した諸条件下で請求項2のcDNAを有する核酸群をハイブリダイズする過程

(c) 複合体形成の量を判定する過程

(d) 前記処理されたサンプル中の複合体形成量を、或る処理されないサンプル中の複合体形成量と比較する過程であって、複合体形成における差が、該分子または化合物の毒性を標示することを特徴とする過程

【請求項22】

或る精製されたタンパク質であって以下から選択されるタンパク質。

(a) SEQ ID NO:1のアミノ酸配列

(b) SEQ ID NO:1の抗原エピトープ、および

(c) SEQ ID NO:1のアミノ酸配列に対して少なくとも90%の配列同一性を有する或るアミノ酸配列

【請求項23】

SEQ ID NO:1のアミノ酸配列を持つ、請求項22に記載の或る精製されたタンパク質。

【請求項24】

請求項22に記載のタンパク質と、或る標識成分とを有する組成物。

【請求項25】

請求項22に記載のタンパク質と、或る医薬用担体とを有する組成物。

【請求項26】

請求項22のタンパク質がその上に固定された或る基板。

【請求項27】

請求項22のタンパク質を有する或るアレイエレメント。

【請求項28】

10

20

30

40

50

SEQ ID NO:1のアミノ酸配列を持つタンパク質の発現をサンプル中に検出する方法であって、

( a ) アッセイを行って或るサンプル中の請求項23のタンパク質の量を判定する過程と、  
( b ) タンパク質の量を標準と比較し、それにより該サンプル中の該タンパク質の発現を検出する過程とを含む方法。

【請求項29】

請求項28の方法であって、アッセイが、抗体アレイ、酵素結合免疫吸着アッセイ ( ELISA )、蛍光活性化細胞選別、二次元ポリアクリルアミドゲル電気泳動およびシンチレーション計数、ラジオイムノアッセイ、並びにウェスタン解析から選択される方法。

【請求項30】

請求項28の方法であって、該サンプルが大腸または胃から得られる方法。

【請求項31】

請求項28の方法であって、該タンパク質が標準と比較して差次的に発現し、大腸癌または胃癌の診断となる方法。

【請求項32】

或るタンパク質を用いて複数の分子および化合物をスクリーニングし、少なくとも1つのリガンドを同定する方法であって、

( a ) 特異結合を許容する条件下で、請求項23に記載のタンパク質を複数の分子および化合物と混合する過程と、

( b ) 特異結合を検出し、それにより該タンパク質に特異結合するリガンドを同定する過程を含む方法。

【請求項33】

請求項32の方法であって、該分子群および化合物群が、アゴニスト、アンタゴニスト、二重特異性分子、DNA分子、小分子薬、イムノグロブリン、インヒビター、擬態物質、多重特異性分子、ペプチド、ペプチド核酸、薬品、タンパク質、およびRNA分子から選択される方法。

【請求項34】

或るタンパク質を用いて、SEQ ID NO:1のアミノ酸配列を持つタンパク質に特異結合する抗体を同定する方法であって、

( a ) 特異結合を許容する条件下で、請求項23に記載のタンパク質と複数の抗体とを接触させる過程と、

( b ) 或る抗体と該タンパク質との特異結合を検出し、それにより該タンパク質に特異結合する抗体を同定する過程を含む方法。

【請求項35】

請求項34の方法であって、該複数の抗体が、ポリクローナル抗体、モノクローナル抗体、キメラ抗体、組換え抗体、ヒト化抗体、一本鎖抗体、Fab断片、F(ab')<sub>2</sub>断片、Fv断片、および抗体-ペプチド融合タンパク質から選択される方法。

【請求項36】

或るタンパク質を用いて或るポリクローナル抗体を調製および精製する方法であって、以下の過程を含む方法。

( a ) 抗体応答を誘発する条件下で、請求項22に記載のタンパク質を用いて或る動物を免疫する過程、

( b ) 動物抗体を単離する過程、

( c ) 該タンパク質を或る基板に接着する過程、

( d ) 該タンパク質への特異結合を許容する条件下で、該基板と単離した抗体群とを接触させる過程、および

( e ) 該抗体群を該タンパク質から解離させ、それにより、精製したポリクローナル抗体群を得る過程

【請求項37】

或るタンパク質を用いて或るモノクローナル抗体を調製する方法であって、以下の過程を

10

20

30

40

50

有する方法。

(a) 抗体応答を誘発する条件下で、請求項22に記載のタンパク質を用いて或る動物を免疫する過程、

(b) 前記動物から、抗体を産出する細胞群を単離する過程、

(c) 不死化した細胞群と前記抗体産出細胞群とを培地内で融合し、モノクローナル抗体を産出するハイブリドーマ細胞群を形成する過程、

(d) 前記ハイブリドーマ細胞を培養する過程、および

(e) 培地から、該タンパク質に特異結合するモノクローナル抗体を単離する過程

【請求項38】

或るタンパク質を用いて癌を診断する方法であって、以下の過程を含む方法。

10

(a) アッセイを行って或るサンプル中の請求項23のタンパク質の発現を定量する過程、および

(b) タンパク質の発現を標準と比較し、それにより癌を診断する過程

【請求項39】

請求項38の方法であって、該サンプルが大腸および胃から得られる方法。

【請求項40】

請求項38の方法であって、発現が大腸癌または胃癌の診断となる方法。

【請求項41】

或る分子または化合物を、アゴニストとしての有効性について試験する方法であって、

20

(a) 請求項23のタンパク質を有する或るサンプルを該分子または化合物に曝す過程と、

(b) 前記サンプルにおいてアゴニスト活性を検出する過程、とを含むことを特徴とする方法。

【請求項42】

或る分子または化合物を、アンタゴニストとしての有効性について試験する方法であって、

(a) 請求項23のタンパク質を有する或るサンプルを或る分子または化合物に曝す過程と

(b) 前記サンプルにおけるアンタゴニスト活性を検出する過程、とを含むことを特徴とする方法。

【請求項43】

30

或る単離された抗体であって、SEQ ID NO:1のアミノ酸配列を持つタンパク質に特異結合することを特徴とする抗体。

【請求項44】

請求項36の方法で産生したポリクローナル抗体。

【請求項45】

請求項37の方法で産生したモノクローナル抗体。

【請求項46】

或る抗体を用いて或るサンプル中の或るタンパク質の発現を検出する方法であって、以下の過程を含む方法。

40

(a) 抗体：タンパク質複合体群の形成を許容する条件下で、請求項43に記載の抗体と或るサンプルとを混合する過程、および

(b) 複合体形成を検出する過程であって、複合体形成が、前記サンプル中の該タンパク質の発現を標示することを特徴とする過程。

【請求項47】

請求項46の方法であって、該サンプルが大腸、肝臓、肺、卵巣、および前立腺から得られる方法。

【請求項48】

請求項46の方法であって、複合体形成が標準と比較され、大腸癌または胃癌の診断となる方法。

【請求項49】

50

或る抗体を用いて或るタンパク質を免疫精製する方法であって、以下の過程を含む方法。

- ( a ) 請求項43に記載の抗体を或る基板に接着する過程
- ( b ) 抗体：タンパク質複合体群の形成を許容する条件下で、該抗体を、タンパク質を有する或るサンプルに曝す過程
- ( c ) 該タンパク質を該複合体から分離する過程、および
- ( d ) 精製されたタンパク質を収集する過程

【請求項50】

請求項43に記載の抗体と、或る標識成分とを有する組成物。

【請求項51】

請求項50の組成物を有する或るキット。

10

【請求項52】

請求項43の抗体を有する或るアレイエレメント。

【請求項53】

請求項43の抗体がその上に固定された或る基板。

【請求項54】

請求項43に記載の抗体と、或る医薬物質とを有する組成物。

【請求項55】

該組成物が凍結乾燥されることを特徴とする、請求項54に記載の組成物。

【請求項56】

或る組成物を用いて或る分子または化合物の効力を算定する方法であって、以下の過程を含む方法。

20

- ( a ) タンパク質を有する或るサンプルを或る分子または化合物で処理する過程
- ( b ) 複合体の形成に適した諸条件下で該サンプル中のタンパク質と請求項54の組成物とを接触させる過程
- ( c ) 複合体形成の量を判定する過程、および
- ( d ) 前記処理されたサンプル中の複合体形成量を、或る処理されないサンプル中の複合体形成量と比較する過程であって、複合体形成における差が、該分子または化合物の効力を標示することを特徴とする過程

【請求項57】

或る組成物を用いて或る分子または化合物の毒性を算定する方法であって、以下の過程を含む方法。

30

- ( a ) タンパク質を有する或るサンプルを或る分子または化合物で処理する過程
- ( b ) 複合体形成に適した諸条件下で、該サンプル中のタンパク質と請求項54の組成物とを接触させる過程
- ( c ) 複合体形成の量を判定する過程、および
- ( d ) 前記処理されたサンプル中の複合体形成量を、或る処理されないサンプル中の複合体形成量と比較する過程であって、複合体形成における差が、該分子または化合物の毒性を標示することを特徴とする過程

【請求項58】

大腸癌または胃癌の治療方法であって、治療介入を要する患者への請求項43の抗体の投与を含む治療方法。

40

【請求項59】

大腸癌または胃癌の治療方法であって、治療介入を要する患者への請求項45の抗体の投与を含む治療方法。

【請求項60】

大腸癌または胃癌の治療方法であって、治療介入を要する患者への請求項54の組成物の投与を含む治療方法。

【請求項61】

或る薬剤を或る細胞に送達する方法であって、以下の過程を含む方法。

- ( a ) 該薬剤を、請求項33に記載の方法で同定した或る二重特異性分子に接着する過程

50

(b) 該二重特異性分子を治療介入を要する或る患者に投与する過程であって、該二重特異性分子がSEQ ID NO:1のアミノ酸配列を持つタンパク質に特異結合し、それにより該薬剤を該細胞に送達することを特徴とする過程

【請求項62】

請求項61の方法であって、該細胞が大腸の上皮細胞である方法。

【請求項63】

請求項23のタンパク質に特異結合する或るアゴニスト。

【請求項64】

請求項63に記載のアゴニストと、或る医薬用担体とを有する組成物。

【請求項65】

請求項23のタンパク質に特異結合する或るアゴニスト。

【請求項66】

請求項65に記載のアンタゴニストと、或る医薬用担体とを有する組成物。

【請求項67】

請求項23のタンパク質に特異結合する或る薬剤。

【請求項68】

請求項67に記載の薬剤と、或る医薬用担体とを有する組成物。

【請求項69】

請求項23のタンパク質に特異結合する或る小分子薬。

【請求項70】

請求項69に記載の小分子薬と、或る医薬用担体とを有する組成物。

【請求項71】

18から30ヌクレオチドの長さのアンチセンス分子であってSEQ ID NO:2の核酸配列を持つ或るポリヌクレオチドの一部に、またはその相補配列に特異結合し、該アンチセンス分子が該ポリヌクレオチドにコードされるタンパク質の発現を阻害するような分子。

【請求項72】

該アンチセンス分子が少なくとも1つの修飾されたヌクレオシド間結合を有する、請求項71に記載のアンチセンス分子。

【請求項73】

該修飾されたヌクレオシド間結合がホスホロチオエート結合である、請求項72に記載のアンチセンス分子。

【請求項74】

該アンチセンス分子が少なくとも1つのヌクレオチド類似体を有する、請求項71に記載のアンチセンス分子。

【請求項75】

該ヌクレオチド類似体が5-メチルシチジンである、請求項72に記載のアンチセンス分子。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、癌において差次的に発現する或る膜貫通タンパク質、それをコードするcDNA、及び該タンパク質に特異結合する或る抗体、並びに、それらを用いた大腸癌または胃癌の診断、病期分類、治療、或いは進行または治療のモニタリングに関する。

【背景技術】

【0002】

アレイ技術と定量PCRは、多数の関連遺伝子または無関係の遺伝子の発現プロファイルを探求する手段を提供する。発現プロファイルを試験するときは、アレイは次のような試験を行うプラットフォームを提供する。即ちどの遺伝子が組織特異的か、ハウスキーピング機能を行うか、シグナル伝達カスケードの一部であるか、又は、特定の遺伝的素因や、条件、疾患、または障害に、特異的に関連する遺伝子であるかの試験である。発現プロファイル作成の応用は特に、疾患の、診断、予後診断、および治療の向上に関する。例えば、

10

20

30

40

50

種々のタイプの癌の患者からの組織に発現される配列群と量の双方を比較しうる。

【0003】

癌と悪性腫瘍は継続的な細胞増殖と細胞死によって特徴付けられ、遺伝学と環境の両方に因果関係がある。癌マーカーは、癌への家族性素因の決定、また様々な癌の早期診断と予後において、非常に重要である。

【0004】

膜貫通タンパク質(TM)、例えば細胞膜を横断するタンパク質は、或る病状の潜在的なマーカーであり治療標的でもある。例えば或る腫瘍細胞に関するならば、多くのTMタンパク質は、細胞内の成長と分化を制御する信号伝達経路群に関する細胞表面受容体として作用する。したがって或る病態では、TMの活性または機能の変調は、疾患経過を妨害しうる。

10

【0005】

大腸直腸癌は、米国で4番目に発生率の高い癌であり癌死の2番目に高い原因で毎年約13万人が新患となり5万5千人が死亡する。大腸癌と直腸癌は多くの環境危険因子を共有しており、両癌は、特有の遺伝的症候群を有する人々に見られる(大腸直腸癌の概説はPotter (1999; J Natl Cancer Institute 91:916-932)を参照)。大腸癌は男女におよそ等しい頻度で発生する唯一の癌である。また米国では大腸癌の診断5年後の生存率は約55%である(Ries他(1990) National Institutes of Health, DHHS Publ No. (NIH) 90-2789)。

【0006】

幾つかの分子経路が大腸癌の発生と関係があり、これらの経路いずれか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高メチル化により失われる可能性がある。発現における変化が大腸癌または大腸癌発生の素因の早期の指標となりうる遺伝子の同定が、特に必要とされている。これらのタンパク質はまた治療標的として利用され、癌の治療に有用な分子が同定されうる。

20

【0007】

大腸癌の素因、発生および進行と関連のある多数の遺伝子が同定されてきた。例えば、ヒトの腫瘍ではDNAメチル化の異常パターンが一貫して発生することがよく知られている。特に大腸癌では、大腸癌に先行する前癌性ポリープ内など、腫瘍進行早期にこれらの変化が発生することが知られる。DNAメチル化を行う酵素であるDNAメチルトランスフェラーゼは、大腸癌または癌に先行する良性ポリープの患者の組織学的に正常な粘膜で著しく増加する。そしてこの増加は大腸腫瘍の進行中継続する(Wafik他(1991) Proc Natl Acad Sci 88:3470-3474)。

30

【0008】

家族性大腸腺腫症(FAP)は大腸癌に先行するまれな常染色体優性症候群であり、大腸腺腫性ポリポーシス(APC)遺伝子の先天性突然変異によって起こる。APC遺伝子は、APC-カテニン-Tcf(T-細胞因子)経路の一部である。この経路の機能障害により、大腸上皮細胞の規則的な複製、接着、遊走が失われ、ポリープの成長につながる。遺伝性非腺腫性大腸癌(HNPCC)は別の先天性常染色体優性症候群であり、大腸癌の早期発症および他癌の発生の傾向により区別する。HNPCCは、DNAミスマッチ修復(MMR)経路の1つ以上の遺伝子の突然変異から発生する。2つのヒトMMR遺伝子MSH2およびMLH1の突然変異は、これまでに同定されたHNPCCファミリーの大部分に見出される。ほとんどすべての大腸癌は、エストロゲン受容体(ER)遺伝子が沈黙した細胞から発生する。ER遺伝子転写の沈黙は年齢に関係しており、ER遺伝子の高メチル化に関連する(Issa他(1994) Nature Genet 7:536-540)。培養した大腸癌細胞に外因性ER遺伝子を導入すると、著しく成長が抑制される。

40

【0009】

大腸癌に関連し、また本病の発生と進行に関連する多数の遺伝的変容が存在することは明らかで、癌発生の早期指標を提供する可能性がある。これらの変容はモニター可能であり、おそらく治療で修正できる。

【0010】

50

或る膜貫通タンパク質、それをコードするcDNA、および該タンパク質に特異結合する抗体の作成法の発見は、大腸癌または胃癌の、診断、病期分類、治療、或いは進行または治療のモニタリングに有用な組成物の提供により、当分野の要望を満たす。

【発明の開示】

【発明の効果】

【0011】

本発明は、癌において差次的に発現する膜貫通タンパク質であるTMDC、それをコードするcDNA、および該タンパク質に特異結合する或る抗体の発見に基づく。これらの分子は、大腸癌または胃癌の、診断、病期分類、治療、或いは進行または治療のモニタリングに有用である。

10

【0012】

本発明は、SEQ ID NO:1のアミノ酸配列を持つ或るタンパク質をコードする或る核酸配列を有する或る単離されたcDNAを提供する。本発明はまた、SEQ ID NO:2の核酸配列、SEQ ID NO:3-10から選択されたSEQ ID NO:2の或る断片、SEQ ID NO:12-16から選択されたSEQ ID NO:2の或る変異体、から選択された或る単離されたcDNAまたはその相補物 (complement) を提供する。本発明は更に、該膜貫通タンパク質をコードするcDNAからなる或るプローブ、該膜貫通タンパク質をコードするcDNAで形質転換される或る細胞、該膜貫通タンパク質をコードするcDNAと或る標識成分とを有する或る組成物、該膜貫通タンパク質をコードするcDNAを有する或るアレイエレメント、及び、該膜貫通タンパク質をコードするcDNAがその上に固定される或る基板、を提供する。

20

【0013】

本発明は、TMDCをコードするcDNAを有する或るベクター、該ベクターを有する或る宿主細胞、及び、該cDNAを用いて該タンパク質を作成する方法を提供し、本法では、該タンパク質をコードする該cDNAを有する該ベクターを有する該宿主細胞を発現に適した条件下で培養し、該タンパク質を該宿主細胞培地から回収する。本発明はまた、TMDCをコードするcDNAを有するベクターを有する或るトランスジェニックの細胞株または生物をも提供する。本発明は更に、該cDNA、その断片、変異体、または相補配列を有する組成物、基板、またはプローブを提供するが、これらは検出、スクリーニング、および精製の方法に使用できる。一態様では、プローブは一本鎖の相補的RNA分子またはDNA分子である。

30

【0014】

本発明は、サンプル中の核酸の差次的発現を検出するためにcDNAを使用する方法を提供する。本法ではプローブを核酸にハイブリダイズし、それによりハイブリダイゼーション複合体を形成して、ハイブリダイゼーション複合体形成を標準と比較する。比較はサンプル中のcDNAの差次的発現を示す。或る態様において検出方法は更にハイブリダイゼーション前にサンプルの核酸の増幅を含む。別の態様では、該cDNAの差次的発現を示す方法を用い、大腸癌または胃癌を診断する。

【0015】

本発明は、或るcDNAを用いて或るライブラリまたは複数の分子あるいは化合物をスクリーニングし該cDNAに特異結合する少なくとも1つのリガンドを同定する方法を提供し、本法では該cDNAと該分子または化合物とを特異結合を許容する条件下で混合し、該cDNAへの特異結合を検出することにより該cDNAに特異結合するリガンドを同定する。一態様では、該分子群あるいは化合物群は、アンチセンス分子、人工染色体作成物、分岐核酸、DNA分子、エンハンサ、ペプチド、ペプチド核酸、タンパク質、RNA分子、リプレッサ、および転写因子から選択される。本発明はまた、或るcDNAを用いて該cDNAに特異結合するリガンドを精製する方法を提供し、本法では該cDNAを或る基板に接着し、該cDNAと或るサンプルとを特異結合を許容する条件下で接触させ、該リガンドを該cDNAから解離させることにより、精製したリガンドを得る。本発明は更に或る分子または化合物の効力または毒性を算定する方法を提供し、本法では、核酸群を有する或るサンプルを該分子または化合物で処理し、該核酸群と或るcDNAとをハイブリダイゼーション複合体が形成される条件下でハイブリダイズし、複合体形成量を判定し、処理済みサンプル中の複合体形成量を無処理サン

40

50

ル中の複合体形成量と比較し、複合体形成の差異が、該分子または化合物の効力または毒性を示す。

【0016】

本発明は、以下からなる群から選択された或る精製されたタンパク質またはその一部を提供する。SEQ ID NO:1のアミノ酸配列、SEQ ID NO:1の抗原エピトープ、及び、SEQ ID NO:1のアミノ酸配列に対して少なくとも90%のアミノ酸配列同一性を持つSEQ ID NO:1の変異体。本発明はまた、該精製タンパク質と或る医薬用キャリアとを有する組成物、該タンパク質と或る標識成分とを有する組成物、該タンパク質が固定される基板、及び、該タンパク質を有する或るアレイエレメントを提供する。本発明は更に、SEQ ID NO:1のアミノ酸配列を持つタンパク質の発現をサンプル中に検出する方法を提供し、本法ではアッセイを行ってサンプル中の該タンパク質の量を判定し、タンパク質量を標準と比較することにより、該サンプル中の該タンパク質の発現を検出する。本発明は更に癌を診断する方法を提供し、本法では、アッセイを行ってサンプル中に発現するタンパク質の量を定量し、発現したタンパク質の量を標準と比較することにより、腫瘍性障害を診断する。一態様では、アッセイは、抗体アレイ、酵素結合免疫吸着アッセイ、蛍光活性化細胞選別、2D-PAGEおよびシンチレーション計数、タンパク質アレイ、ラジオイムノアッセイ、並びにウェスタン解析から選択される。第二の態様では、該サンプルが大腸および胃の組織から選択される。第三の態様では、癌は大腸癌または胃癌である。

10

【0017】

本発明は、或るタンパク質を用いて或るライブラリまたは複数の分子あるいは化合物をスクリーニングし少なくとも1つのリガンドを同定する方法を提供し、本法では該タンパク質と該分子または化合物とを特異結合を許容する条件下で混合し、特異結合を検出することにより該タンパク質に特異結合するリガンドを同定する。一態様では、該分子群および化合物群が、アゴニスト、アンタゴニスト、二重特異性分子、DNA分子、小分子薬、イムノグロブリン、インヒビター、擬態物質、多重特異性分子、ペプチド、ペプチド核酸、医薬品、タンパク質、およびRNA分子から選択される。別の態様では、該リガンドを用いて腫瘍性障害患者を治療する。本発明はまた、SEQ ID NO:1のアミノ酸配列を有するタンパク質に特異結合する治療用抗体を提供する。本発明はまた、SEQ ID NO:1のアミノ酸配列を有するタンパク質に特異結合するアンタゴニストを提供する。本発明はまた、SEQ ID NO:1のアミノ酸配列を有するタンパク質に特異結合する小分子薬を提供する。本発明はまた、リガンドがアゴニストまたはアンタゴニストとして有効であるか試験する方法を提供し、本法では、該タンパク質を有するサンプルを該分子または化合物に曝す過程と、サンプル中のアゴニストまたはアンタゴニスト活性を検出する過程とを含む。

20

30

【0018】

本発明は、或るタンパク質を用いて該タンパク質に特異結合する抗体を同定するため複数の抗体をスクリーニングする方法を提供し、本法では複数の抗体と該タンパク質とを抗体:タンパク質複合体を形成する条件下で接触させ、該抗体を該抗体:タンパク質複合体から解離させることにより、該タンパク質に特異結合する抗体を得る。一態様では、該抗体が、無傷イムノグロブリン分子、ポリクローナル抗体、モノクローナル抗体、二重特異性分子、多重特異性分子、キメラ抗体、組換え抗体、ヒト化抗体、一本鎖抗体、Fab断片、F(ab')<sub>2</sub>断片、Fv断片、および抗体-ペプチド融合タンパク質から選択される。本発明は、或るタンパク質に特異結合する精製された抗体群を提供する。

40

【0019】

本発明はまた、或るタンパク質を用いて、該タンパク質に特異結合するポリクローナル抗体およびモノクローナル抗体を調製および精製する方法を提供する。ポリクローナル抗体を調製する方法では、抗体応答を誘発する条件下で動物をタンパク質で免疫し、動物抗体を単離し、該タンパク質を或る基板に接着し、該基板と単離した抗体とを該タンパク質への特異結合を許容する条件下で接触させ、該抗体を該タンパク質から解離させることにより、精製したポリクローナル抗体を得る。モノクローナル抗体を調製する方法では、抗体応答を誘発する条件下で動物を或るタンパク質で免疫し、抗体産生細胞を動物から単離し

50

、該抗体産生細胞と不死化細胞とを培地で融合してモノクローナル抗体産生ハイブリドーマ細胞を形成し、該ハイブリドーマ細胞を培養し、モノクローナル抗体を培地から単離する。

【0020】

本発明はまた、或る抗体を用いてサンプル中の或るタンパク質の発現を検出する方法を提供し、本法では該抗体とサンプルとを抗体：タンパク質複合体を形成する条件下で混合し、複合体形成を検出し、複合体形成はサンプル中の該タンパク質の発現を示す。一態様では、該サンプルが大腸または胃の組織から選択される。第二の態様では、複合体形成が標準と比較され、大腸癌または胃癌の診断となる。

【0021】

本発明は、或るタンパク質の免疫精製方法を提供し、本法では或る抗体を或る基板に接着し、該抗体を、抗体：タンパク質複合体を形成する条件下で、タンパク質を有する或るサンプルに曝し、該タンパク質を該複合体から解離させ、精製したタンパク質を収集する。本発明はまた、該タンパク質に特異結合する或る抗体と或る標識成分または医薬品とを有する或る組成物、該組成物を有する或るキット、該抗体を有する或るアレイエレメント、該抗体が固定される或る基板を提供する。本発明は更に或る抗体を用いて或る分子または化合物の効力を算定する方法を提供し、本法では、タンパク質を有する或るサンプルを或る分子または化合物で処理し、該サンプル中のタンパク質と該抗体とを複合体が形成される条件下で接触させ、複合体形成量を判定し、処理済みサンプル中の複合体形成量を無処理サンプル中の複合体形成量と比較し、複合体形成の差異が、該分子または化合物の効力を示す。

【0022】

本発明は大腸癌の或る治療法を提供し、本法は、治療介入を要する或る被検者に、該タンパク質に特異結合する或る治療用抗体、該タンパク質に特異結合する或る二重特異性分子、該タンパク質に特異結合する或る多重特異性分子、または、或る抗体と或る医薬品とを有する或る組成物を投与する。本発明はまた、或る医薬物質または治療物質を或る細胞に送達する方法を提供し、本法では、該医薬物質または治療物質を該タンパク質に特異結合する或る二重特異性分子と付着させ、治療介入を要する或る被検者に該二重特異性分子を投与し、該二重特異性分子が該医薬物質または治療物質を該細胞に送達する。一態様では、該細胞は大腸の上皮細胞である。

【0023】

本発明は、該タンパク質に特異結合する或るアゴニストおよび、該アゴニストと或る医薬用キャリアとを有する或る組成物を提供する。本発明は又、該タンパク質に特異結合する或るアンタゴニストおよび、該アンタゴニストと或る医薬用キャリアとを有する或る組成物を提供する。本発明はまた、該タンパク質に特異結合する或る医薬品または小分子薬を提供する。

【0024】

本発明は、18から30ヌクレオチドの長さのアンチセンス分子であってSEQ ID NO:2の核酸配列を持つ或るポリヌクレオチドの一部に、またはその相補物に特異結合し、該アンチセンス分子が、該ポリヌクレオチドにコードされるタンパク質の発現を阻害するような分子を提供する。

【0025】

本発明は又、少なくとも1つの修飾されたヌクレオシド間結合または少なくとも1つのヌクレオチド類似体を持つ或るアンチセンス分子を提供する。本発明はまた、該修飾されたヌクレオシド間結合がホスホロチオエート結合である、また該修飾された核酸塩基が5-メチルシトシンである分子を提供する。

【0026】

本発明は、内在性ポリヌクレオチドの発現を攪乱するために異種マーカ-遺伝子を哺乳動物のゲノムDNAに挿入する方法を提供する。

さらに本発明は、哺乳動物モデル系を構築するためにcDNAを使用する方法を提供し、こ

10

20

30

40

50

ではSEQ ID NO:2-16から選択されたcDNAを持つベクターを作製し、ベクターをES細胞に形質転換し、形質転換したES細胞を選択し、形質転換したES細胞を哺乳動物胚盤胞に微量注入し、それによってキメラ胚盤胞を形成し、キメラ胚盤胞を偽妊娠メスに移植し、前記メスが生殖細胞系に該cDNAを持つキメラ子孫を産み、そしてホモ接合性哺乳動物モデル系を構築するため該キメラ哺乳動物を交配する。

【発明を実施するための最良の形態】

【0027】

(本発明の記載について)

本発明が、説明した特定の装置、材料及び方法に限定されるものではないことを理解されたい。また、ここで使用する専門用語は特定の実施態様を説明する目的で用いたものであり、特許請求の範囲にのみ限定される本発明の範囲を限定することを意図したものではないことも併せて理解されたい。

10

【0028】

請求の範囲および明細書中で用いる単数形の「或る」および「その(この)」の表記は、文脈から明らかにそうでないとされる場合を除いて複数ものを指す場合もある。例えば「或る宿主細胞」と記されている場合には当業者に公知のそのような宿主細胞が複数あることもある。

【0029】

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

20

【0030】

(定義)

「抗体」とは、無傷イムノグロブリン分子、ポリクローナル抗体、モノクローナル抗体、キメラ抗体、組換え抗体、ヒト化抗体、一本鎖抗体、Fab断片、F(ab')<sub>2</sub>断片、Fv断片、および抗体-ペプチド融合タンパク質を指す。

【0031】

「抗原決定基」は、抗原エピトープまたは免疫原エピトープを指し、これはオリゴペプチド、ペプチド、またはタンパク質の構造的特徴または領域であって該タンパク質に特異結合する抗体の形成を誘導するものを指す。生物活性は、免疫原性に必須ではない。

30

【0032】

「アレイ」は、或る基板上の、少なくとも2つのcDNA、タンパク質、または抗体の、秩序ある配置を指す。少なくとも1つのcDNA、タンパク質、または抗体は対照または標準であり、他のcDNA、タンパク質、または抗体が、診断または治療上の関心事である。基板上の少なくとも2個、最大で約40,000のcDNA、タンパク質、または抗体の配置が、各々標識される複合体(各cDNAと少なくとも1個の核酸との間、各タンパク質と少なくとも1個のリガンドまたは抗体との間、または各抗体と該抗体が特異結合する少なくとも1個のタンパク質との間に形成される複合体)の、サイズと信号強度が個々に識別できるよう保証する。

40

【0033】

「二重特異性分子」は2種の結合特異性を持ち、2種の分子、または或る分子の2箇所部位に同時に結合できる。同様に、「多重特異性分子」は多数(3つ以上)の別の標的に結合でき、その1つは免疫細胞の表面の分子である。抗体は二重特異性分子または多重特異性分子として機能でき、また、これらの一部と成りうる。

【0034】

「TMDC」は、SEQ ID NO:1のアミノ酸配列に対して厳密または高度に相同な(>85%)膜貫通タンパク質であって、天然、合成、半合成或いは組換え体など任意の出所からの、任意の種(例えばウシ、ヒツジ、ブタ、ネズミ、ウマ及び好適にはヒト種)から得られる膜貫

50

通タンパク質を指す。

【0035】

配列表 (Sequence Listing) のcDNAの「相補物」は、核酸分子であって、その全長にわたって完全に相補的であり、高ストリンジェンシー条件下で或る核酸分子とハイブリダイズする分子である。

【0036】

「cDNA」とは、或る単離されたポリヌクレオチド、核酸分子、またはその何れかの断片であって、約400～約12000ヌクレオチドを持つものを指す。cDNAは起源が組換えまたは合成のものがあり、一本鎖または二本鎖があり、コードおよび非コードの3'または5'配列を示すことがあり、また一般にイントロンを欠く。

10

【0037】

用語「タンパク質をコードするcDNA」は、核酸であって、その配列が、公知の分析で同定される保存された領域、モチーフまたはドメインをコードする配列群と密にアラインするものを指す。この分析としては、BLAST (Basic Local Alignment Search Tool; Altschul (1993) J Mol Evol 36:290300; Altschul 他 (1990) J Mol Biol 215:403-410) およびBLAST2 (Altschul 他 (1997) Nucleic Acids Res 25:3389-3402) があり、これらは保存された領域内の同一性を提示する。Brenner 他 (1998; Proc Natl Acad Sci 95:6073-6078) はBLASTを分析して構造的相同体を配列同一性によって同定する性能について調べ、30%の同一性が少なくとも150の残基の配列アラインメントでの信頼できる閾値であり40%が少なくとも70の残基のアラインメントでの合理的閾値であることを発見した (Brenner, 6076ページ, 列2)。

20

【0038】

「組成物」は、ポリヌクレオチドおよび標識成分を指しまた、精製したタンパク質と医薬用キャリアまたは異種成分、標識成分あるいは精製成分を指し、抗体と標識成分または医薬品、なども指す。

【0039】

「誘導体」とは化学修飾されたcDNAあるいはタンパク質を指す。cDNAの誘導体化には、クエオシン (queosine) などの非従来型塩基の或いはヒポキサンチンなどの類似体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。cDNAまたはタンパク質の誘導体化にはまた、アセチル基、アシル基、アルキル基、アミノ基、ホルミル基、またはモルホリノ基による水素の置換 (例えば5-メチルシトシン) を含むうる。誘導体分子は天然分子の生物活性を保持するが、長い寿命または増強した活性を与える。

30

【0040】

「差次的発現」とは、サンプル中の転写されたメッセンジャーRNAあるいは翻訳されたタンパク質の存在の有無や、量の少なくとも二倍の変化により検出される、増加または上方調節あるいは減少または下方調節発現を指す。

【0041】

「障害 (disorder)」は、病状、疾患、または症候群であって、そこでTMDCまたはTMDCをコードするmRNAが差次的に発現するものを指し、これには大腸癌および胃癌がある。

【0042】

「発現プロファイル」は、或るサンプルでの遺伝子発現の表現である。核酸発現プロファイルを作成するには、シーケンシング、ハイブリダイゼーション、または増幅 (定量PCR) の技術と、サンプルからのmRNAまたはcDNAを用いる。タンパク質発現プロファイルは、時間の遅延はあるが核酸発現プロファイルの鏡像をなし、これには抗体またはタンパク質アレイ、酵素結合免疫吸着アッセイ、蛍光活性化細胞選別、空間的固定たとえば2D-PAGE、および、ラジオイムノアッセイたとえば放射標識およびシンチレーションカウンタでの定量、並びにウェスタン解析を用いてサンプル中のタンパク質発現を検出する。核酸、タンパク質、または抗体は溶液でまたは基板に付着させて使用でき、その検出は公知の方法と標識成分に基づく。発現プロファイルを評価する方法にはまた、電子的ノーザン分析、「guilt-by-association」 (連帯責任により有罪) 法、および転写物イメージングが

40

50

ある。上記いずれかの方法で作成した発現プロファイルは、正常または患部組織を用いて作成した発現プロファイルと対比できる。注目すべきは、mRNAとタンパク質の発現の対応が、Zweiger (2001, Transducing the Genome. McGraw-Hill, San Francisco, CA) および Glavas 他 (2001; T cell activation upregulates cyclic nucleotide phosphodiesterases 8A1 and 7A3, Proc Natl Acad Sci 98:6319-6342) などで述べられている。

【0043】

「断片」とは長さが約50から約5000塩基対の連続したヌクレオチドの鎖を指す。断片は、関連する核酸分子を同定するためにPCRまたはハイブリダイゼーション技術で、またリガンドをスクリーニングするために結合アッセイで使用される可能性がある。このようなりガンドは、複製、転写または翻訳を調節する治療薬として有用である。

10

【0044】

「Guilt-by-association」(GBA)法は、或る特定の疾患、調節経路、細胞内区画、細胞タイプ、組織タイプ、または種に関連するcDNAまたはタンパク質の同定を、既知のマーカ—または治療薬との、高度に有意な共発現で行う方法である。

【0045】

「ハイブリダイゼーション複合体」は、1つの分子のプリン群が5'-A-G-T-C-3'塩基対と3'-T-C-A-G-5'のように相補分子のピリミジン群と水素結合する時、或るcDNAと或るサンプルの核酸の間で形成される。ハイブリダイゼーション条件、相補性の度合、およびヌクレオチド類似体の使用は、ハイブリダイゼーション反応の効率とストリンジェンシーに影響する。

20

【0046】

配列に用いられる用語「同一」とは、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る2つ以上の配列間のヌクレオチドまたは残基の一致の定量化(通常、百分率)を指し、アルゴリズムは例えばSmith-Waterman alignment (SmithおよびWaterman (1981) J Mol Biol 147:195-197), CLUSTALW (Thompson 他 (1994) Nucleic Acids Res 22:4673-4680), またはBLAST2 (Altschul (1997, 前出)がある。BLAST2は、アラインメントを最適化するために配列の1つにギャップを挿入するまた配列をより有意に比較できる、標準化された再現性のある方法で使用される可能性がある。「類似性」には同じアルゴリズムを使うが、残基の保存的置換を考慮する。タンパク質では類似性が同一性を上回り、理由はパリンのロイシンまたはイソロイシンとの置換が、報告される百分率の計算でカウントされるからである。保存的とみなされる置換は当分野で公知である。

30

【0047】

「単離された」または「精製された」は、いずれかの分子または化合物が、その自然環境から分離され、天然に会合する他の成分が約60%~90%無いものを指す。

【0048】

「標識成分」は、任意のレポーター分子、例えば放射性核種、酵素、蛍光剤、化学発光剤、色原体、基板、補助因子、阻害因子、または磁気粒子であって、ポリヌクレオチド、タンパク質、または抗体に、付着または組込みできる分子を指す。可視の標識および色素としては、限定はされないが、アントシアニン、グルクロニダーゼ、ビオチン、BIODIPY、クーマシーブルー、Cy3とCy5、4,6-ジアミジノ-2-フェニルインドール(DAPI)、ジゴキシゲニン、フルオレセイン、FITC、金、緑色蛍光タンパク、リサミン(lissamine)、ルシフェラーゼ、フィコエリトリン、ローダミン、spyro red、銀、ストレプトアビジンなどがある。放射性マーカーには、水素、ヨウ素、リン、硫黄などの放射性形態が含まれる。

40

【0049】

「リガンド」とは、或るポリヌクレオチドに或いは或るタンパク質のエピトープに特異結合する、任意の薬剤、分子または化合物を指す。このようなりガンドはポリヌクレオチドまたはタンパク質の活性を安定化あるいは調節する。そしてミネラル、補助因子、核酸、タンパク質、糖質、脂肪、脂質など、無機および/または有機物質から構成されている可能性がある。

50

## 【0050】

「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。これはハイブリダイゼーションまたは増幅技術、或いは複製、転写または翻訳の調節で使用される可能性がある。等価な用語に、アンプリコン（単位複製配列）、アンプリマー、プライマー、およびオリゴマーがある。

## 【0051】

「医薬品（pharmaceutical agent）」は抗体の場合もあり、アンチセンス分子、二重特異性分子、多重特異性分子、ペプチド、タンパク質、放射性核種、小分子薬、細胞特異的または細胞障害性の薬物たとえばアブリン、アクチノミオシンD、シスプラチン、crotin、ドキシソルピシン、5-フルオロウラシル、メトトレキサート、リシン（ricin）、ピンクリスチン、ピンブラスチン、または、これらエレメントの任意の併用でもありうる。

10

## 【0052】

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

## 【0053】

「プローブ」とは、サンプル中の少なくとも1つの核酸にハイブリダイズするcDNAを指す。標的が一本鎖である場合、プローブは相補的一本鎖である。プローブは、サザン法、ノーザン法、*in situ*、ドットプロット法、アレイなどの技術を含め、ハイブリダイゼーション反応で、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できる。

20

## 【0054】

「タンパク質」とは、ポリペプチドあるいはその任意の部分の指す。或るタンパク質の「一部/部分」とは、ある長さのアミノ酸配列であって、該タンパク質の、少なくとも1つの生物活性、PFAMまたはPRINTS分析で同定される或るドメイン、または、PROTEANプログラム（DNASTAR, Madison WI）のKyte-Doolittleアルゴリズムで同定される或る抗原決定基を保持する配列を指す。「オリゴペプチド」は、抗体を産生する融合タンパク質の一部として使用される、約5残基から約15残基までのアミノ酸配列である。

## 【0055】

用語「サンプル」は、その最も広い意味で用いられ、核酸、タンパク質、及び抗体を有するものがある。サンプルは、体液たとえば腹水、血液、脳脊髄液、リンパ、精液、唾液、尿などや、細胞製剤の可溶分画、または、その中で細胞が成長した培地のアリコットや、細胞から単離または抽出した染色体や細胞内小器官（オルガネラ）や膜や、溶液中に存在するまたは基板に結合したゲノムDNA、RNA、cDNAや、細胞、組織、組織生検、組織プリント、口腔粘膜細胞、皮膚、毛髪、毛包などを有し得る。

30

## 【0056】

「特異結合」は、二分子間の正確な相互作用であって、それらの構造、特にそれらの分子側基に依存する作用をさす。例えば、調節タンパク質のDNA分子主溝への挿入（インターカレーション）、あるいは、タンパク質のエピトープとアゴニスト、アンタゴニストまたは抗体との間の結合がある。

40

## 【0057】

「基板」は、任意の剛性あるいは半剛性の支持物であってポリヌクレオチド、タンパク質、または抗体が結合するものを指し、磁気または非磁気ビーズ、毛細管または他のチューブ、チップ、ファイバー、フィルタ、ゲル、膜、プレート、ポリマー、スライド、ウェハ、および微小粒子であって種々の表面形状たとえばチャネル、カラム、ピン、細孔、トレンチ、およびウェルなどを持つものが含まれる。

## 【0058】

「転写イメージ（transcript image: TI）」は、特定の組織内、特定時の、遺伝子転写活性のプロファイルである。TIは或るESTデータベースのcDNAライブラリ群の中で発現され

50

るポリヌクレオチド群の相対的存在度のアセスメントを提示し、これは米国特許第5,840,484号に記載され、参照により開示に含まれる。

【0059】

「変異体」は、分子であって、或るタンパク質または其れをコードするポリヌクレオチド群の、認識されるパリエーションをいう。スプライス変異体はBLASTスコアにより決定でき、スコアは少なくとも100、最も好ましいのは少なくとも400である。対立遺伝子変異体はcDNAに対して高一致率を有し、100塩基につき約3塩基が異なりうる。「1塩基多型」(SNP)とは、置換、挿入または欠失による、単一塩基での変異を指す。変異は保存的(プリンからプリン)あるいは非保存的な(プリンからピリミジン)可能性があり、コード化されたアミノ酸またはその2次、3次、4次構造の変異となる可能性もならない可能性もある。

10

【0060】

(発明)

本発明は、癌において差次的に発現する膜貫通タンパク質、該タンパク質をコードする或るcDNA、および該タンパク質に特異結合する或る抗体の発見に基づく。該タンパク質、またはその各部分、該cDNA、またはその断片群、および該抗体を、直接に、または組成物として用いて、大腸癌または胃癌を診断、病期分類、治療、或いは其の進行または治療をモニターできる。

【0061】

本発明のTMDCをコードする核酸は初め、或る大腸腫瘍ライブラリ(COLNTUT03)からのInc yte Clone 1929823の中に、或るヌクレオチドおよび/またはアミノ酸配列アラインメント用コンピュータ検索を用いて同定された。SEQ ID NO:2の由来は以下のオーバーラップおよび/または伸長した核酸配列群(SEQ ID NO:3-10)およびその関連cDNAライブラリ群Inc yte Clone 1929823H1、1929823T6、および1341151F6(COLNTUT03)、7703595H1(UTRET UE01)、8146316H1(MIXDTME01)、3274531H1、ショットガン配列であるSCCA02331V1およびSCCA04417V、並びに、ゲノム配列g2951946\_010(SEQ ID NO:11)である。

20

【0062】

一実施態様において本発明はSEQ ID NO:1のアミノ酸配列を持つ或るタンパク質を含み、これは図(Figure) 1Aから1Hに示す。TMDCは760アミノ酸の長さであり、7つの潜在的N-グリコシル化部位をアミノ酸残基N16、N77、N221、N264、N342、N350、およびN567に持つ。TMDCは、1つの潜在的サイクリックAMP/サイクリックGMP依存タンパク質キナーゼ燐酸化部位をT237に、11ヶ所の潜在的カゼインキナーゼ燐酸化部位をS46、S48、S75、S97、T115、T129、S174、T241、S474、S634およびS752に、6つの潜在的タンパク質キナーゼC燐酸化部位をS59、T115、T148、T188、S640およびS749に、また1つの潜在的チロシンキナーゼ燐酸化部位をY536に持つ。HMMR分析では、以下の9つの膜貫通ドメインがある。TM-1、アミノ酸残基210-230; TM-2、アミノ酸残基281-299; TM-3、アミノ酸残基372-392; TM-4、アミノ酸残基447-467; TM-5、アミノ酸残基487-507; TM-6、アミノ酸残基540-562; TM-7、アミノ酸残基586-610; TM-8、アミノ酸残基654-672; およびTM-9、アミノ酸残基956-974である。SEQ ID NO:1の有用な抗原エピトープ群は、ほぼアミノ酸残基S110~ほぼR150、ほぼF230~ほぼG270、ほぼV330~ほぼF370、および、ほぼC420~ほぼI450に及ぶ。膜貫通タンパク質腫瘍抗原に特異結合する抗体は、癌、特に大腸癌または胃癌を同定する診断アッセイに有用である。

30

40

【0063】

図2はTMDCの疎水性プロットであり、種々の膜貫通領域を疎水性領域(プロットのY軸上の負の値)として示す。

【0064】

図3は、種々の正常な成体組織をTMDC発現につきTAQMAN解析で分析した結果である。最も顕著なTMDC発現は精巣、脂肪組織、乳房、十二指腸、および結腸に見られ、TMDCが比較的限局的な正常組織分布を持つことを示す。しかし精巣での高発現は、内部対照である2ミクログロブリンでの、正常より高い発現に関連した。

50

## 【 0 0 6 5 】

表 (Table) 1は、LIFESEQ Goldデータベース (Incyte Genomics) 内のcDNAライブラリ群のノーザン分析による、組織カテゴリ群でのTMDCの発現を示す。結果は、消化系でのTMDCの最高の存在量 (Abundance: 見られた転写物の総数) を示す。上記の表1と図3の結果に見られる差異は、おそらく、LIFESEQデータベースのcDNAライブラリ群における胎児組織と患部組織の高出現率を反映する。

## 【 0 0 6 6 】

また表2は、消化系の内、TMDCを過剰発現するcDNAライブラリ群 (転写物/ライブラリの比が1より大きい群) が患部組織であり、大腸腫瘍、FAP、炎症性の腸、および胃の腫瘍を含むことを示す。特に注目すべきは大腸腫瘍 (COLNTUT03) を同じドナーからの正常大腸組織 (COLNNOT16) (TMDC発現は検出不能であった) と比べた場合のTMDCの過剰発現と、胃の腫瘍であるSTOMTUP02での過剰発現でありSTOMTUP02はTMDCを発現する消化組織の内での最高の存在量を示した。

## 【 0 0 6 7 】

図4は、QPCR分析 (Applied Biosystems) を用いて正常大腸組織と比較した、大腸癌組織サンプル群でのTMDC発現を示す。結果は、試験した9の内8サンプルの大腸腫瘍でのTMDCの発現の増加を示す。結果は、癌組織と正常組織とで少なくとも1.2倍の発現差が見られたときに、有意とした。

## 【 0 0 6 8 】

図5は同様に、QPCR分析を用いて非腫瘍化大腸細胞株LS123と比較した、種々のヒト大腸腫瘍細胞株でのTMDC発現を示す。TMDCは、試験した8の内6の大腸腫瘍細胞株、LS174、HCT116、Caco2、HT29、COL0205およびSW620で過剰発現する。

## 【 0 0 6 9 】

表3は、マイクロアレイ分析で、大腸癌組織群と正常大腸組織とでのTMDCの発現を比較した結果である。結果は、試験した14の内2名の患者でのTMDCの発現の増加を示す。差次的発現 (列1) は、癌組織と正常組織とで少なくとも1.5倍の発現差が見られたときに、有意とした。図3のQPCRで分析したサンプル群と表1とを比較した相対発現値の差の部分的原因は、おそらく、マイクロアレイ分析に比してのQPCR分析の感度とダイナミックレンジの大きさである。

## 【 0 0 7 0 】

TMDCをコードするcDNAの哺乳類変異体群の同定には、BLAST2で初期値パラメータ群とZOOSEQデータベース群 (Incyte Genomics) を用いた。これらの好適な変異体は、以下の表で示されるように約84%から90%の同一性を有する。列1は変異体cDNAのSEQ IDvar、列2は変異体cDNAのクローン番号、列3は種、列4はヒトcDNAとの一致率、列5は変異体cDNAのヒトcDNAへのアラインメントである。

SEQ ID <sub>var</sub>	cDNA <sub>var</sub>	種	一致率	N <sub>tH</sub> アラインメント
12	701294553H1	ラット	85%	474-654
13	701600294H1	ラット	88%	2927-2994
14	2016808H1	マウス	89%	2939-3052
15	239780_Mm.1	マウス	84%	714-1615
16	703528478J1	イヌ	90%	2927-2990

## 【 0 0 7 1 】

遺伝暗号の縮重により、TMDCをコードする多数のcDNAが作り出され、中には、既知のいか

なる天然遺伝子のcDNAとも極小の類似性しか持たない配列もあることは、当業者には理解されよう。従って本発明は、可能コドン選択に基づく組合せの選択によって産出し得るような、ありとあらゆる可能性のあるcDNA変異を企図する。これらの組合せは、天然TMDCをコードするポリヌクレオチドに適用される標準的なトリプレット遺伝暗号を基に作られる。また、このような全ての変異が明確に開示されていると考慮されたい。

#### 【0072】

SEQ ID NO:2-16のcDNAは、SEQ ID NO:2とサンプル中の関連分子の中で同定、区別するための、ハイブリダイゼーション、増幅、およびスクリーニング技術で使用可能である。哺乳類cDNAであるSEQ ID NO:12-16を用いてトランスジェニックの細胞株または生物を産生できるが、これらはヒトの大腸癌または胃癌のモデル系であり、それらについて治療処置の毒性と有効性を試験できる。本発明の、cDNA、タンパク質、抗体、並びに、cDNAとタンパク質とを用いて同定される分子および化合物を使用して、毒性研究、臨床試験、および被検者/患者治療プロファイルを、実施またモニターすることが可能である。

10

#### 【0073】

本発明の性質決定と使用

##### cDNAライブラリ

ここに開示する或る特定の実施態様では、mRNAを当業者に周知の方法を使用して哺乳動物の細胞と組織から単離し、cDNAライブラリを準備するために使用する。Incyte cDNA群を、「実施例」のように準備した哺乳類cDNAライブラリ群から単離した。コンセンサス配列は1つのクローン挿入中に在るか化学的にアセンブリされ、これはIncyte cDNA群ならびに伸長配列および/またはショットガン配列など、シークエンスした断片群からの電子アセンブリに基づく。コンピュータープログラム例えばPHRAP (P Green, University of Washington, Seattle WA) およびAUTOASSEMBLERアプリケーション (ABI) を用いて配列アセンブリし、これらは実施例5に記載する。5' および3' 配列の検証後、TMDCをコードするIncyteクローン1929823F6を、研究開発用試薬に指定した。

20

#### 【0074】

##### 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該分野で周知であり、本発明の何れの実施態様も、核酸シークエンシング法を用いて実行可能である。これらの方法に用いる酵素は例えばDNAポリメラーゼIのクレノウ断片、SEQUENASE, Taq DNAポリメラーゼ、および熱安定性T7 DNAポリメラーゼ (Amersham Biosciences (APB), Piscataway NJ)、または、ポリメラーゼ群と校正エキソヌクレアーゼ群 (Invitrogen, Carlsbad CA) の併用がある。配列の準備はMICROLAB 2200システム (Hamilton, Reno NV) およびDNA ENGINEサーマルサイクラー (MJ Research, Watertown MA) およびABI CATALYST 3700サーマルサイクラー (Applied Biosystems) などの装置を、シークエンシングはPRISM 3700, 377または373 DNAシークエンシングシステム (ABI)、或いはMEGABACE 1000 DNAシークエンシングシステム (APB) を用いて自動化する。

30

#### 【0075】

配列表に在るcDNA群の核酸配列の準備にはこうした自動化法を用い、この配列群には時折、シークエンスエラーとNで示す未同定ヌクレオチドとを含みうるが、これはcDNAをシークエンスした時点での最新技術を反映する。ベクター、リンカーおよびポリ(A)配列をマスクするには、BLASTと、動的プログラミングと、隣接ジヌクレオチド分析 (dinucleotide nearest neighbor analysis) とに基づく、アルゴリズムとプログラムとを用いた。N群とSNP群を検証するには、cDNAを再シークエンスするか、該N群またはSNP、が起きるエリアにオーバーラップする多数の配列を比較するアルゴリズムを用いる。この両技術は当業者に周知であり使用もされる。配列の分析には種々のアルゴリズムを使用でき、それらはAusubel他 (1997; Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7) およびMeyers (1995; Molecular Biology and Biotechnology, Wiley VCH, New York NY, 856-853ページ) に記載がある。

40

#### 【0076】

50

ショットガン・シーケンシングを、当該の特定クローン化インサートの配列を完成するために使用することもできる。ショットガン・ストラテジーには、無作為に元来のインサートを多様なサイズのセグメントに切断すること、これらの断片をベクターにクローニングすることが含まれる。これらの断片は、元来のインサートの配列全体がわかるまで、オーバーラップする末端を使用してシーケンスされ再アセンブリされる。ショットガン・シーケンシング法は当分野で公知であり、本法は、当該のcDNAに隣接する代表的領域群から選択される、熱安定性DNAポリメラーゼ、易熱性DNAポリメラーゼ、およびプライマーを使用する。不完全アセンブリした配列の同一性の点検には種々のアルゴリズムまたはプログラム例えばCONSED (Gordon (1998) Genome Res 8:195-202)を用い、これらは当該分野で公知である。コンタミネートする配列たとえばベクターまたはキメラ配列は除去でき、欠失した配列は復旧可能で、アセンブリした完全な配列を完成できる。

10

【0077】

#### 核酸配列の伸長

本発明の配列群の伸長には、種々のPCRベースの、当分野で既知の方法を利用できる。例えばXL-PCR kit (ABI)、入れ子プライマー群、およびcDNAまたはゲノムDNAライブラリを用いて核酸配列を伸長できる。全てのPCRベースの方法では、ソフトウェア、例えばOLIGOプライマー分析ソフトウェア (Molecular Biology Insights, Cascade CO)を用いて、長さが約22~30ヌクレオチド、GC含量が約50%以上、温度約55~約68で標的分子にアニーリングするように、プライマー群を設計し得る。配列を伸長して調節エレメント群を回収する際は、cDNAでなくゲノムライブラリ群を用いる。

20

【0078】

#### ハイブリダイゼーション

cDNAとその断片は、様々な目的のハイブリダイゼーション技術において使用することが可能である。プローブの設計または誘導は、固有領域 (例えば5'調節領域) から、また、非保存領域 (すなわち、該タンパク質の保存触媒ドメインをコードするヌクレオチド群の5'または3') から可能で、プローブは、TMDCをコードする天然分子、対立遺伝子変異体、または近縁分子を同定するプロトコルに用い得る。プローブはDNAまたはRNAの場合があり、一本鎖の場合があり、核酸配列SEQ ID NOs:2-9の何れかに対して少なくとも50%の配列同一性を持つ。ハイブリダイゼーションプローブを産生するには、オリゴ標識化、ニックトランスレーション、末端標識化 (エンドラベリング)、または、レポーター分子の存在下のPCR増幅を用い得る。cDNAまたはその断片を有するベクターを用いてmRNAプローブを *in vitro* で産生でき、このためには或るRNAポリメラーゼと標識したヌクレオチド群を付加する。これらの手順には、APBが提供するようなキットを用い得る。

30

【0079】

ハイブリダイゼーションのストリンジェンシーは、プローブのG+C含量、塩濃度、および温度により決定する。特にストリンジェンシーは、塩の濃度の減少あるいはハイブリダイゼーション温度の上昇により増強することができる。ハイブリダイゼーションは、5xSSCに1%のドデシル硫酸ナトリウム (SDS) で60などのバッファーでの低ストリンジェンシーで実施することができ、そこでは、いくらかのミスマッチを有する核酸配列間でハイブリダイゼーション複合体の形成を可能にする。より高いストリンジェンシーでは、続いて45 (中間ストリンジェンシー) または68 (高ストリンジェンシー) で、0.1% SDSを有する0.2xSSCなどのバッファーで洗浄を実施する。高ストリンジェンシーでは、ハイブリダイゼーション複合体は核酸が完全に相補的な場合のみ安定を保持するであろう。数種の膜系ハイブリダイゼーションでは、約35%~約50%のホルムアミドをハイブリダイゼーション溶液に添加し、ハイブリダイゼーションが成される温度を低下できる。背景信号を低減するには、界面活性剤たとえばSarkosylまたはTRITON X-100 (Sigma-Aldrich)、およびブロッキング剤たとえば変性サケ精子DNAを用い得る。構成成分の選択とハイブリダイゼーションの条件とは当業者には公知である。またAusubel (前出) およびSambrook他 (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainville NYに概説されている。

40

50

## 【0080】

アレイを当該分野で周知の方法を用いて調製また分析することができる。オリゴヌクレオチドまたはcDNAは、同時に多数の遺伝子の発現レベルをモニターするために、または遺伝変異体、突然変異及び一塩基多型を同定するために、ハイブリダイゼーションプローブあるいは標的として使用され得る。アレイを用いることで、遺伝子機能を決定し、症状、疾患または障害の遺伝的根拠を理解し、症状、疾患または障害を診断し、治療薬を開発し活性をモニターすることができる（例えば米国特許第5,474,796号；Schena他（1996）Proc Natl Acad Sci 93:10614-10619；Heller他（1997）Proc Natl Acad Sci 94:2150-2155；米国特許第5,605,662号参照）。

## 【0081】

ハイブリダイゼーションプローブはまた、天然のゲノム配列をマッピングするのにも有用である。ハイブリダイゼーションプローブは、特定の染色体、染色体の特定領域、または人工染色体作成物にハイブリダイズされうる。こうした作成物としては、ヒト人工染色体、酵母人工染色体、細菌人工染色体、細菌P1作成物、または、単一染色体から作製したライブラリ群のcDNAがある。

## 【0082】

## QPCR

QPCR法は核酸分子の定量法であり、PCR増幅時に発生する蛍光信号の検出に基づく（Gibson 他（1996）Genome Res 6:995-1001；Heid 他（1996）Genome Res 6:986-994）。増幅に用いる装置は例えばPRISM 7700検出システム（ABI）であり、これは96穴サーマルサイクラーがレーザーおよび電荷結合素子（CCD）光学系に接続される。QPCRを行うには、PCR反応を、二重標識したプローブの存在下で行う。プローブは標準の順行と逆行のPCRプライマー間でアニールするよう設計され、5'末端で蛍光原性レポーター色素（たとえば6-カルボキシフルオレセイン（6-FAM））によって、また3'末端でクエンチャー分子（たとえば6-カルボキシテトラメチルローダミン（TAMRA））によって標識される。プローブが無傷な内は、3'クエンチャーが、5'レポーターによる蛍光を消す。しかし各プライマー伸長周期の間、アニール済プローブは、Taqポリメラーゼの内在性の5' → 3'ヌクレアーゼ活性の結果、分解する（Holland 他（1991）Proc Natl Acad Sci 88:7276-7280）。この分解はレポーターをクエンチャーから分離し、蛍光は数秒ごとにCCDで検出される。核酸の開始コピー数が高いほど、早期に蛍光の増加が見られる。サイクル閾値（ $C_T$ ）はPCR産物が一定の検出閾値を超えるサイクル数であり、機器のソフトウェアが判定する。 $C_T$ は鋳型のコピー数に逆比例するので、サンプル中の核酸分子の相対または絶対的初濃度の計算に用い得る。2種の分子の相対濃度を計算するには、各々の $C_T$ 値を判定して行える（比較 $C_T$ 法）。或いは核酸分子の絶対濃度を計算するには、既知濃度の或るハウスキーピング分子を用いて標準曲線を作成して行える。 $C_T$ 値を計算し、標準曲線を準備し、開始コピー数を判定する過程には、SEQUENCE DETECTOR 1.7ソフトウェア（ABI）を用いる。

## 【0083】

## 発現

TMDCをコードする多数のcDNAの何れか1つをベクターにクローン化でき、該タンパク質または其の各部分を宿主細胞内で発現するために用い得る。核酸配列は、新規な制限酵素部位の作製、グリコシル化パターンの改変、コドン選択性の変更のために、DNAシャッフリング（USPN 5,830,721）、部位特異的変異誘発などの方法により操作することが可能であり、これにより、特定宿主における発現の増大、スプライス変異体の生成、半減期の延長などがなされる。発現ベクターは、特定の宿主での効率のために選択された多様な出所からの、転写及び翻訳調節エレメント（プロモータ、エンハンサ、特定の開始シグナル、ポリアダニル化された3'配列）を持つことが可能である。ベクター、cDNA、調節エレメントは、当分野で公知でありSambrook（前出4、8、16、17章）に記載されている、*in vitro*組換えDNA技術、合成技術、及び/または*in vivo*遺伝的組換え技術を利用して結合する。

## 【0084】

種々の宿主系を、発現ベクターで形質転換可能である。これらには、限定するものではな

10

20

30

40

50

いが組換えバクテリオファージ、プラスミドまたはコスミドDNA発現ベクターで形質転換させた細菌や、酵母発現ベクターで形質転換させた酵母や、バキュロウイルス発現ベクターで形質転換させた昆虫細胞系や、ウイルス性および/または細菌性エレメントを有する発現ベクターで形質転換させた植物細胞系がある(前出Ausubel、ユニット16)。哺乳動物細胞系においては、アデノウイルス転写/翻訳複合体を利用し得る。配列をウイルスのゲノムのE1あるいはE3領域にライゲーションさせた後に、この感染性ウイルスが、宿主細胞を形質転換しそこで該タンパク質を発現するために用いられる。ラウス肉腫ウイルスエンハンサーあるいは、SV40またはEBVをベースにしたベクターを用いて、タンパク質を高レベル発現させることもできる。

#### 【0085】

核酸配列の慣例的なクローニング、サブクローニング、および増殖には、多機能のpBLUES CRIPトベクター(Stratagene, La Jolla CA)またはpSPORT1プラスミド(Invitrogen)を用いることができる。或る核酸配列をこれらベクターのマルチクローニング部位に導入するとlacZ遺伝子が攪乱され、形質転換された細菌のための比色スクリーニング法が可能となる。更にこれらのベクターは、クローニングされた配列におけるin vitro転写、ジデオキシのシーケンシング、ヘルパーファージによる一本鎖のレスキュー、入れ子欠失の生成にも有用であろう。

#### 【0086】

組換えタンパク質の長期産生のために、同じ或いは別のベクターの上の選択可能マーカ―または可視マーカ―遺伝子と共に、ベクターは安定的に細胞株に形質転換されうる。形質転換後、細胞は強化培地で約1~2日間増殖させ、次に選択培地に移す。選択可能マーカ―や、代謝拮抗物質、抗生物質或いは除草剤耐性の遺伝子は関連する選択剤への耐性を与え、導入された配列の発現に成功するような細胞の成長及び回収が可能となる。選択培地上の生存にあるいは可視マーカ―の発現により同定された耐性クローンは、培養技術を用いて増殖可能である。可視マーカ―はまた、導入された遺伝子により発現したタンパク質の量を推定するためにも用いられる。所望のcDNAを有する宿主細胞の検証は、DNA-DNAまたはDNA-RNAハイブリダイゼーションあるいはPCR増幅に基づく。

#### 【0087】

宿主細胞は、所望の形で組換えタンパク質を修飾する能力について、選択され得る。そのような修飾には、アセチル化、カルボキシル化、グリコシル化、リン酸化、脂質化、アシル化及びその他がある。「プレプロ」形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、折りたたみ、および/または活性を特定することも可能である。翻訳後の活性のための、特定の細胞機構および特徴的な機序を持つ、種々の宿主細胞を、組換えタンパク質の妥当な修飾およびプロセッシングを確実にするように選択し得る。

#### 【0088】

##### 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ベクターに組み込まれる異種成分には、グルタチオンSトランスフェラーゼ(GST)、6xHis、FLAG、MYC及びその他がある。GSTと6-Hisを精製するには、それぞれ固定化グルタチオンおよび金属キレート樹脂など、親和性マトリクスを用いる。FLAGおよびMYCを精製するには、モノクローナルおよびポリクローナル抗体を用いる。精製後の分離を容易にするために、タンパク質と異種成分の間に位置するベクターに、タンパク質分解切断部位をコードする配列を含めることもできる。組換えタンパク質の発現および精製方法は、Ausubel(前出、ユニット16)に記載されている。

#### 【0089】

##### タンパク質の同定

高速液体クロマトグラフィと質量分析(MS)を用いてタンパク質の迅速な同定を可能にする数種の技術が開発されている。タンパク質を有するサンプルを発端に、本法では、1)タンパク質を二次元ゲル電気泳動(2-DE)で分離し、2)選択したタンパク質をゲルから切除しプロテアーゼで消化して一組のペプチドを産生し、3)ペプチドを質量分析してペプチドイオン質量とスペクトルパターン情報を引き出す。MS情報を用いてタンパク質を同

10

20

30

40

50

定するには、この情報をタンパク質データベースの情報と比較する (Shevenko 他 (1996) Proc Natl Acad Sci 93:14440-14445)。蛋白質の2DEでの分離には、等電点分離法 (IEF) を1次元目、その後SDS-PAGEを2次元目に用いる。IEFでは、固定したpH勾配ストリップが、分離の再現性と分解能の増進に有用である。別の技術を用いて、極めて塩基性、疎水性、または高分子量のタンパク質の分解能を改善しうる。分離したタンパク質の検出には、MSに適合する染色液または色素たとえば銀染色、クーマシーブルー、またはspyro red (Molecular Probes, Eugene OR) を利用できる。ゲルはPVDF膜にブロットしてウェスタン分析でき、光学スキャンにSTORMスキャナー (APB) を用いて、コンピュータで読み取れる出力を産出可能で、それをパターン認識ソフトウェア (例えばMELANIE (GeneBio, Geneva, Switzerland)) で分析する。本ソフトウェアは個々のスポットを注釈するが、これは、一意の識別子を割当て、各々のX、Y座標、分子量、等電点、および信号強度を計算して行う。当該の個々のスポット (例えば差次的に発現したタンパク質を示すスポット) が切除され、部位特異的プロテアーゼ (例えばトリプシンまたはキモトリプシン) の単独または併用で蛋白分解消化され、一組の小ペプチド、好適には1~2 kDaの範囲のものが産生される。消化の前にサンプルを還元剤およびアルキル化剤で処理でき、消化後にペプチドを液体クロマトグラフィまたはキャピラリー電気泳動で分離し、MSで分析する。

10

#### 【0090】

MSはサンプルの各成分をガス状イオンに変換し、イオンを質量電荷比に基づき分離し、相対存在度を判定する。ペプチドマスフィンガープリント分析では、MALDI-TOF (Matrix Assisted Laser Desorption/Ionization-Time of Flight、マトリックス支援レーザー脱離イオン化飛行時間型質量分析)、ESI (Electrospray Ionization、エレクトロスプレーイオン化質量分析)、およびTOF-TOF (Time of Flight/Time of Flight) 装置を用いて、一組の高精度なペプチド質量を判定する。分析プログラム例えばTURBOSEQUENTソフトウェア (Finnigan, San Jose CA) を用いてMSデータを、既知または予測したタンパク質に由来する理論的MSデータのデータベースと比較する。最少で3ペプチド質量の一致を用いて、信頼できるタンパク同定を行う。更なる情報が同定に必要であれば、タンデムMSを用いて個々のペプチドの情報を引き出せる。タンデムMSでは、第1ステージのMSで個々のペプチド質量を判定する。次に、選択したペプチドイオンを断片化するため衝突誘起解離 (CID) などの技術を用い、一連のイオンを産生する。結果の断片化イオン群を第2ラウンドのMSで分析し、それらのスペクトルパターンを用いてアミノ酸配列の短いストレッチを判定できる (Dancik 他 (1999) J Comput Biol 6:327-342)。

20

30

#### 【0091】

該タンパク質がデータベースにあれば、ペプチド質量と断片化データの組合せは、該タンパク質の算出されたMWおよびpIと共に、通常、明白な同定を産生することとなる。一致が見られない場合、タンパク配列を得るには、当該分野で公知の直接化学シーケンシング法を用い得る (Creighton (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY参照)。

#### 【0092】

##### ペプチドの化学合成

タンパク質または其の各部分を産生するには、組換え法に限らず当該分野で公知の化学法も用いうる。固相ペプチド合成はバッチ式または連続フロープロセスで実行でき、これは逐次的に、アミノ保護および側鎖保護したアミノ酸残基を不溶性ポリマー支持体に、或るリンカー基を介して付加する。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基を、サポートレジンを形成するためスチレン-ジビニルベンゼン共重合体に付着する。これらアミノ酸残基は、酸不安定Boc (t-ブチルオキシカルボニル) または塩基不安定Fmoc (9-フルオレニルメトキシカルボニル) でN-保護される。保護アミノ酸のカルボキシル基はリンカー基のアミンに結合させ、この残基を固相サポートレジンにアンカーする。トリフルオロ酢酸またはピペリジンが、それぞれBocまたはFmocの場合に保護基を除去するために用いられる。各付加的アミノ酸を、カップリング剤か前もって活性化されたアミノ酸誘導体を用いてアンカー済み残基に付加し、次いでレジンを洗浄する。全長ペ

40

50

プチドを、逐次脱保護、誘導体化アミノ酸の結合、ジクロロメタンまた/あるいはN, N-ジメチルホルムアミドでの洗浄により合成する。前記ペプチドを、ペプチド酸またはアミドを産生するためにペプチドカルボキシ末端とリンカー基の間で切断する (Novabiochem 1997/98 Catalog and Peptide Synthesis Handbook, San Diego CAS1-S20ページ)。自動合成はまた、431Aペプチドシンセサイザ (ABI) 等の装置で成し得る。タンパク質或いはその部分は、分取用高速液体クロマトグラフを用いて精製可能である。またその組成はアミノ酸分析かシーケンシングにより確認することができる (Creighton (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY)。

#### 【0093】

##### 抗体

抗体すなわちイムノグロブリン (Ig) は免疫応答の成分であり、B細胞の表面で発現されるか、または、B細胞によって分泌されて血液循環に入る。プロトタイプの抗体は、ジスルフィド結合によって連結された2つの同じポリペプチド重鎖 (H鎖) と2つの同じポリペプチド軽鎖 (L鎖) からなる四量体であり、外来抗原に結合し中和する。抗体はH鎖に基づいてIgA、IgD、IgE、IgGまたはIgMに分類される。最も一般的なクラスであるIgGは四量体であるが、他のクラスのもの基本構造の変異体か、または多量体である。

#### 【0094】

抗体は、其の2つの機能ドメインについて記述される。抗原認識を媒介するのは抗体のFab (抗原結合フラグメント) 領域であり、エフェクター機能を媒介するのはFc (結晶可能フラグメント) 領域である。抗体の抗原への結合は、抗原の、食作用白血球たとえばマクロファージおよび好中球による破壊を誘起する。これらの細胞は表面Fc受容体を発現し、これが抗体のFc領域に特異結合し、食細胞に抗体結合抗原を破壊させる。Fc受容体は1回膜貫通糖タンパクで約350アミノ酸を持ち、細胞外部分は通常、2または3のIgドメインを持つ (Sears他 (1990) *J Immunol* 144:371-378)。

#### 【0095】

##### 抗体の調製とスクリーニング

種々の宿主たとえばマウス、ラット、ウサギ、ヤギ、ラマ、ラクダ、ヒト細胞株などが、抗原決定基の注入によって免疫化され得る。アジュバント例えばフロイント、ミネラルゲル、および界面活性物質 (例えばリゾレシチン)、プルロニックポリオール (pluronic polyols)、ポリアニオン、ペプチド、油性乳剤、スカシガイのヘモシアニン (KLH; Sigma -Aldrich)、ジニトロフェノールなどを用いて免疫応答を増進しうる。ヒトでは、BCG (カルメット ゲラン杆菌) およびコリネバクテリウム パルバム (Corynebacterium parvum) が応答を増進する。抗原決定基は、オリゴペプチド、ペプチド、またはタンパク質でありうる。抗原決定基の量が免疫化を反復するものであれば、高親和性の特異的ポリクローナル抗体を得ることができる (Klinman および Press (1975) *Transplant Rev* 24:41-83)。内因性タンパク質の一部に同一な約5~約15のアミノ酸をもちうるオリゴペプチドをKLHなどのタンパク質と融合し、該キメラ分子への抗体を産生しうる。

#### 【0096】

モノクローナル抗体は、培地内の連続継代性細胞株によって抗体を産生する任意の技術を用いて調製することが可能である。この技術としては、ハイブリドーマ法、ヒトB細胞ハイブリドーマ法、およびEBVハイブリドーマ法がある (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; およびCole 他 (1984) *Mol Cell Biol* 62:109-120)。

#### 【0097】

キメラ抗体を作製するには、ヒト抗体遺伝子にマウス抗体遺伝子をスプライシングするなどの技術が、好適な抗原特異性および生物活性を備える分子を得るために用いられる (Morrison他 (1984) *Proc Natl Acad Sci* 81:6851-6855; Neuberger 他 (1984) *Nature* 312:604-608; およびTakeda他 (1985) *Nature* 314:452-454)。別法では、抗体産生のために記載された技術を適用し、当分野で既知の方法を用いて、特異的一本鎖抗体を産生しうる。関連した特異性を有するガイディオタイプ組成が異なるような抗体群を、ランダムな

10

20

30

40

50

組合せの免疫グロブリンライブラリからチェーンシャッフリングによって産生することもできる (Burton (1991) Proc Natl Acad Sci 88:10134-10137)。或る抗原決定基に対する特異結合部位を持つ抗体断片をも産生し得る。例えば、限定するものではないが、このような断片には、抗体分子のペプシン消化によって作製される  $F(ab')_2$  断片と、 $F(ab')_2$  断片のジスルフィド架橋を還元することによって作製される Fab断片とがある。あるいは、Fab発現ライブラリを作製することによって、所望の特異性を持つモノクローナル Fab断片を迅速且つ容易に同定できる (Huse他 (1989) Science 246:1275-1281)。

#### 【0098】

抗体類の産生はまた、リンパ球集団における産生の誘導によって、或いは非常に特異的な結合試薬の免疫グロブリンのライブラリ類またはパネル類のスクリーニングによっても行え、これは Orlandi 他 (1989; Proc Natl Acad Sci 86:3833-3837) または Winter 他 (1991; Nature 349:293-299) に開示がある。タンパク質をファージミドまたは Bリンパ球イムノグロブリンのライブラリのスクリーニングアッセイに用いて、所望の特異性を有する抗体を同定し得る。確立された特異性を有するポリクローナル抗体またはモノクローナル抗体の何れかを用いる、イムノアッセイまたは競合結合に関する数々のプロトコルは、当該分野で公知である。

#### 【0099】

##### 抗体特異性

ラジオイムノアッセイ技術と共に Scatchard 分析などの様々な方法を用いて、或るタンパク質に対する特定抗体の親和性を算定しうる。親和性を結合定数  $K_a$  で表すが、この  $K_a$  は、平衡状態の下でタンパク-抗体複合体のモル濃度を遊離抗体と遊離抗原のモル濃度で除して得られる値である。ポリクローナル抗体類は多様な抗原決定基に対する親和性が不均一であり、或るポリクローナル抗体製剤に関して判定した  $K_a$  は、抗体群の平均の親和性または結合活性を表す。モノクローナル抗体は特定の抗原決定基に特異的で、モノクローナル抗体の或る製剤について判定した  $K_a$  は、親和性の真の測定値を表す。 $K_a$  値が約  $10^9 \sim 10^{12}$  L/mol の高親和性抗体製剤は、蛋白-抗体複合体が激しい操作に耐えなければならないイムノアッセイに通常用いる。 $K_a$  値が約  $10^6 \sim 10^7$  L/mol の低親和性抗体製剤は、該タンパク質が抗体から最終的に (好ましくは活性型で) 解離する必要がある免疫精製 (immunopurification) および類似の処理に用いるのが好ましい (Catty (1988) *Antibodies, Volume 1: A Practical Approach*, IRL Press, Washington DC; Liddell および Crayer (1991) *A Practical Guide to Monoclonal Antibodies*, John Wiley & Sons, New York NY)。

#### 【0100】

ポリクローナル抗体製剤の抗体価および結合活性を更に評価して、後に使う或る適用例に対するこのような製剤の品質および適合性を決定することができる。例えば、約 5~10mg/ml の特異抗体を含むポリクローナル抗体製剤は一般に、蛋白-抗体複合体を沈殿させなければならない処理に用いる。抗体作製や、抗体の特異性、力価、および結合力の評価の手順、並びに、様々な適用例における抗体の品質と使用に対する指針については、Catty (前出) および Ausubel (前出) 11.1-11.31 ページに記載がある。

#### 【0101】

##### 細胞の形質転換アッセイ

細胞の癌化、すなわち正常な細胞の癌細胞への転換は、高度に複雑で遺伝的に多様な過程である。しかし、この過程に伴う、細胞生理の数種の変容は、*in vitro* の細胞ベースのシステムまたは *in vivo* の動物モデルを用いてアッセイできる。既知の変容としては、成長シグナルに関する獲得性の自足性、成長阻害シグナルへの不感性、無制限の複製能、アポトーシスの回避、持続する血管形成、および細胞浸潤ならびに転移がある (Hanahan および Weinberg (2000) Cell 100:57-70 参照)。こうしたアッセイを用いて例えば、細胞内の TMDc などの遺伝子の過剰発現の、細胞癌化への効果をアセスできる。

#### 【0102】

(診断)

10

20

30

40

50

TMDC、それをコードするcDNA、またはTMDCに特異結合する或る抗体と、以下の少なくとも1つのアッセイを用いて検出するTMDCの差次的発現を用いて、大腸癌または胃癌を診断できる。

#### 【0103】

##### アッセイのための分子の標識化

多岐にわたるレポーター分子及び結合技術が当業者に知られており、様々な核酸アッセイ、アミノ酸アッセイおよび抗体アッセイにこれらの技術を用い得る。標識した分子の合成を成すのに用いるキットとしては、Promega (Madison WI) またはAPBが供給するような、標識したヌクレオチド例えば<sup>32</sup>P-dCTP (APB)、Cy3-dCTPまたはCy5-dCTP (Qiagen-Operon, Alameda CA)、或いはアミノ酸たとえば<sup>35</sup>S-メチオニン (APB) の組み込み用キットがある。ヌクレオチドとアミノ酸を直接に標識するには、種々の物質たとえば蛍光剤、化学発光剤、色原体などを用い、BIODIPYまたはFITC (Molecular Probes) 等の試薬を用いた、分子内に在る、アミン、チオール、および他の基への化学接合によって行う。

10

#### 【0104】

##### 核酸アッセイ

cDNA、断片群、オリゴヌクレオチド、相補RNA、およびペプチド核酸 (PNA) を用いて、差次的な遺伝子発現を検出および定量し、障害を診断できる。同様に、TMDCに特異結合する抗体を用いて、該タンパク質を定量し得る。こうした差次的発現を伴う障害としては、大腸癌または胃癌がある。診断アッセイは、差次的遺伝子発現を検出するために患者の生物学的サンプル中の遺伝子発現を標準サンプルと比較する、ハイブリダイゼーションまたは増幅技術を用い得る。この比較のための定性方法または定量方法は、当分野で公知である。

20

#### 【0105】

##### 発現プロファイル

発現プロファイルには複数のcDNAまたはタンパク質の発現を有し、この測定には、サンプルを有する標準アッセイを用いる。本発明のcDNA、タンパク質、または抗体を用いてアレイ上のエレメントとし、発現プロファイルを作成できる。一実施態様では、該アレイを用いて、疾患を診断、またはその進行を監視する。

#### 【0106】

例えば、cDNAあるいはプローブは標準的な方法で標識化し、ハイブリダイゼーション複合体の形成に好適な条件下で患者の生物学的サンプルに加える。インキュベーション期間後、サンプルを洗浄し、ハイブリダイゼーション複合体と関連する標識 (またはシグナル) の量を定量して標準値と比較する。患者サンプル中の複合体形成が正常または疾患の標準に比して変容すれば、差次的発現は障害の存在を示す。

30

#### 【0107】

差次的発現を確立する標準を提供するために、正常と疾患の発現プロファイルを確立する。これは、ハイブリダイゼーション発生に好適な条件下で、cDNAと、動物或いはヒトの何れかの正常な被検者から取ったサンプルを混合することにより達成される。精製された配列を既知量用いて行った実験から得た値を正常な対象から得た値と比較することにより、標準ハイブリダイゼーション複合体を定量することができる。このようにして得た標準値は、特定の症状、疾患または障害と診断された患者からのサンプルから得た値と比較することができる。標準値から特定の障害に伴う値への偏差を用いて、その障害を診断または病期分類する。

40

#### 【0108】

遺伝子発現のパターンでの変化を分析することにより、疾患を早期に、患者に症状が出る前に診断できる。本発明を用いて予後診断を策定し、治療レジメンを設計できる。本発明はまた、治療の有効性のモニタリングに利用できる。既知の副作用を持つ治療については、該アレイを用いて治療レジメンを改善する。治療成功の指標となる、遺伝子発現パターンでの変化を起こす用量が確立される。望ましくない副作用の発症を伴う発現パターンは回避される。この手法は、患者が不十分な改善または副作用を示すのを待って治療のコー

50

スを変更するよりも、高感度かつ迅速でありうる。

【0109】

別の実施態様では、ヒト疾患を模倣する動物モデルを用いて、特定の病状、疾患、または障害、或いは、該病状、疾患または障害の治療に伴う発現プロファイルの特徴づけできる。新規の治療レジメンをこれらの動物モデルで試験でき、これにはアレイを用い、発現プロファイルを確立した後、長期間フォローする。またアレイを動物モデルから採取した細胞培地または組織と共に用いて多数の候補薬物分子を迅速にスクリーニングでき、これは既知の治療薬の発現に似た発現プロファイルを産出する分子を探すもので、同じ発現プロファイルを持つ分子がおそらく似た治療効果を持つことを見込む。このように、本発明は、薬物の分子作用の機序を迅速に判定する手段を提供する。

10

【0110】

このようなアッセイを、動物実験または臨床試験における特定の治療処置レジメンの効果を評価するため、或いは個々の患者の治療をモニターするために用いることもできる。症状の存在を一旦確定して治療プロトコルを開始すると、診断アッセイを定期的に繰り返して、患者の発現レベルが正常な被検者に観察されるレベルに近づき始めたかどうかを判定し得る。経時的なアッセイから得られた結果を用いて、数日から数年の期間にわたる治療の効力を示し得る。

【0111】

タンパク質アッセイ

特異的ポリクローナル抗体または特異的モノクローナル抗体を用いてタンパク質発現の尺度としての複合体形成の検出及び計測を行うための免疫学的方法は、当分野で既知である。このような技術の例としては、抗体アレイ、酵素結合免疫吸着アッセイ、蛍光活性化細胞選別、2D-PAGEとシンチレーション計数、蛋白アレイ、ラジオイムノアッセイ、およびウェスタン分析がある。こうしたイムノアッセイには通常、タンパク質と其の特異抗体との複合体形成の測定を含む。これらのアッセイと其の精製した標識化標準に対する定量化とは当分野で公知である（前出のAusubel、unit 10.1-10.6）。2つの非干渉エピトープに反応するモノクローナル抗体群を用いた2部位モノクローナル系イムノアッセイ（two-site, monoclonal-based immunoassay）が好ましいが、競合結合試験を用いることもできる（Pound（1998）Immunochemical Protocols, Humana Press, Totowa NJ）。

20

【0112】

これらの方法はまた、差次的なタンパク質発現を示す疾患の診断に有用である。正常或いは標準的なタンパク質発現の値は、複合体形成に適した条件下で、正常な哺乳動物またはヒト被検者から採取した体液または細胞抽出物と或るタンパク質に対する特異抗体とを混合することによって確定する。正常組織および患部組織での標準複合体形成値は、種々の方法、多くは測光法で確定する。次に、被験サンプルで発現する複合体形成を標準値と比較する。正常標準から、また罹患標準への偏差は疾患の診断または予後のためのパラメータを提供し、罹患標準からまた正常標準への偏差は治療効力の評価に用い得る。

30

【0113】

最近、抗体アレイによって、組換え抗体の高処理スクリーニングの技術の開発が可能となった。こうした方法ではロボットを用いて抗体遺伝子含有細菌を採取および格子配列し、フィルタ系ELISAを用いて抗体断片を発現するクローンをスクリーニングし同定する。液体ハンドリングが排除されクローンはマスターストックから配列されるので、同じ抗体を何度もスポットでき、複数の抗原に対して同時にスクリーニングできる。抗体アレイは、差次的に発現するタンパク質の同定で有用性が高い（de Wildt他（2000）Nature Biotechnology 18:989-94参照）。

40

【0114】

（治療）

TMDC（SEQ ID NO:1）および他の膜貫通タンパク質の領域の間には、化学的及び構造的類似性、特に膜貫通ドメインが存在する。また、差次的発現は、大腸癌および胃癌と高度に関連する。TMDCは明らかに大腸癌または胃癌で役割を果たす。

50

## 【0115】

一実施態様では、該タンパク質の発現または活性の低下を望む場合、抗体、アンタゴニスト、インヒビター、医薬品、または、これら分子の1つ以上を有する組成を、こうした治療を要する被検者に送達しうる。こうした送達は当該分野で公知の各法で行え、該タンパク質に特異結合する抗体による送達もありうる。治療用には、モノクローナル抗体を用いて活性部位を遮断し、二量体形成を阻害し、アポトーシスを誘起するなどする。

## 【0116】

別の実施態様では、該タンパク質の発現または活性の増加を望む場合、該タンパク質、アゴニスト、エンハンサ、医薬品、または、これら分子の1つ以上を有する組成を、こうした治療を要する被検者に送達しうる。こうした送達は当該分野で公知の各法で行え、該タンパク質を特異的に標的とする抗体による医薬品の送達もありうる。

10

## 【0117】

任意のcDNA、相補分子、または其の断片、タンパク質または其の各部分、これらの核酸分子を送達または該タンパク質を発現するベクター、治療抗体、及び、該cDNAまたはタンパク質に結合するリガンドを、別の治療薬と併用投与できる。併用療法で用いる治療薬は、当業者が従来 of 医薬原理に従って選択し得る。治療薬の併用により、各薬剤の、より低い用量で、特定の障害の治療に影響する相乗効果をもたらし得る。

## 【0118】

核酸を用いる遺伝子発現の修飾

TMDCをコードする遺伝子のコントロール領域、5'、3'、または他の調節領域に相補的またはアンチセンスの分子(DNA、RNA、またはPNA)を設計して遺伝子発現をモディフィケーションし得る。転写開始を阻害するように設計されたオリゴヌクレオチドが好ましい。同様に阻害は、ポリメラーゼ、転写因子または調節分子の結合を阻害する三重らせん塩基対合を用いて達成することができる(Gee他In: HuberおよびCarr (1994) Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, 163-177ページ)。相補分子はまた、mRNAとリボソームの間の結合を阻止することによって翻訳を遮断するように設計することができる。或いは、ライブラリまたは複数のcDNAを、調節性非翻訳配列に特異結合するものを同定するためにスクリーニングすることが可能である。

20

## 【0119】

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

30

## 【0120】

本発明の相補的な核酸及びリボザイムは、in vitroまたはin vivoで組換え発現によりあるいは固相フォスフォアミダイト化学合成を使用して調製することが可能である。さらに、分子の5'末端、3'末端、あるいはその両方においてフランキング配列を付加したり、分子の主鎖内においてホスホジエステラーゼ結合ではなくホスホロチオネート(phosphorothioate)または2'-O-メチルを使用したりすることにより、細胞内の安定性を高め半減期を長くするためにRNA分子を修飾することができる。修飾はPNAの産出に固有のものであり、他の核酸分子に拡張することができる。非従来型塩基、例えばイノシン、クエオシン(queosine)、ワイプトシン(wybutosine)を含めるか、アデニン、シチジン、グアニン、チミン、およびウリジンをアセチル-、メチル-、チオ基で修飾すると、分子は内因性エンドヌクレアーゼへの耐性を増す。

40

## 【0121】

cDNA治療

50

本発明のcDNA群を遺伝子治療に用い得る。cDNAはex vivoで標的細胞たとえば骨髄の細胞に送達できる。安定した組込みと転写および/または翻訳を確認したら、骨髄を被検者に再導入できる。該cDNAがコードするタンパク質の発現は、正常配列の突然変異、内因性標的タンパクの減少または喪失、或いは、内因性または突然変異タンパクの過剰発現を伴う障害を修正しうる。或いはcDNAをin vivoで、ベクター例えばレトロウイルス、アデノウイルス、アデノ随伴ウイルス、単純ヘルペスウイルス、および細菌プラスミドを用いて送達できる。非ウイルス性遺伝子送達法としては、カチオン性リポソーム、ポリリジン抱合体、人工ウイルスエンベロープ、および、DNAの直接注入がある (Anderson (1998) Nature 392:25-30; Dachs他 (1997) Oncol Res 9:313-325; Chu他 (1998) J Mol Med 76 (3-4) :184-192; Weiss他 (1999) Cell Mol Life Sci 55 (3) :334-358; Agrawal (1996) Antisense Therapeutics, Humana Press, Totowa NJ; およびAugust他 (1997) Gene Therapy (Advances in Pharmacology, Vol. 40), Academic Press, San Diego CA)。

#### 【0122】

##### モノクローナル抗体治療

抗体、特に、特定のタンパク質、酵素、または受容体に特異結合し其の過剰発現を阻止するモノクローナル抗体が現在、治療に使われている。初めて広く受容された治療抗体は、HERCEPTIN (Trastuzumab, Genentech, S. San Francisco CA) およびGLEEVEC (メシル酸イマチニブ, Novartis Pharmaceuticals, East Hanover NJ) である。HERCEPTINはヒト化抗体であり、HER2陽性転移性乳癌の治療で承認された。これは、過剰発現したHER2タンパクに結合し機能をブロックするよう設計される。GLEEVECの適応は、フィラデルフィア染色体陽性 (Ph+) 慢性骨髄性白血病 (CML) の芽球クリーゼ (急性転化期) または慢性期でインターフェロン治療無効後の患者の治療である。GLEEVECの第二の適応は、KIT (CD117) 陽性で切除不能および/または転移性の悪性胃腸間質腫瘍の患者の治療である。他のモノクローナル抗体群は臨床試験の種々の段階にあり、適応は例えば前立腺癌、リンパ腫、黒色腫、肺炎球菌感染、リウマチ様関節炎、乾癬、全身性エリテマトーデスなどがある。

#### 【0123】

##### スクリーニング及び精製アッセイ

TMDCをコードするcDNAを用いて、或るライブラリまたは複数の分子あるいは化合物の特異結合親和性をスクリーニングできる。ライブラリは、アンチセンス分子、人工染色体作成物、分岐核酸分子、DNA分子、ペプチド、ペプチド核酸、タンパク質 (たとえば転写因子、エンハンサまたはリプレッサ)、RNA分子、リボザイム、および他のリガンドであり、内因性遺伝子の、活性、複製、転写、または翻訳を制御するものでありうる。アッセイには、或るポリヌクレオチドと、或るライブラリまたは複数の分子あるいは化合物を、特異結合を許容する条件下で混合する過程と、特異結合を検出して該cDNAに特異結合する少なくとも1つの分子を同定する過程がある。

#### 【0124】

一実施態様において、本発明のcDNAを複数の精製された分子または化合物と共にインキュベートすることが可能であり、ゲル遅延測定法 (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の方法によって結合活性を判定することが可能である。別の実施態様において、cDNAは生検および/または培養の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。cDNAと核抽出物の分子あるいは化合物の間の特異結合は、最初ゲルシフト法により決定される。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、ゲル遅延アッセイにスーパーシフトが起きる。

#### 【0125】

別の実施態様では、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。一実施態様では、cDNAは重合樹脂またはゲル上の臭化シアン基と化学反応させる。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、収集されうる。

【0126】

更なる実施態様では、該タンパク質あるいはその一部分を、サンプルからのリガンドを精製するために用いる。タンパク質を用いてリガンドを精製する方法では、タンパク質とサンプルを、特異結合できる条件で混合し、タンパク質とリガンドの特異結合を検出し、結合タンパクを回収し、カオトロピック剤を用いてタンパク質を精製リガンドから分離することになる。

【0127】

或る好適な実施態様では、TMDCを用いて、複数の分子または化合物を、種々のスクリーニングアッセイの何れかでスクリーニングできる。そのようなスクリーニングで用いるタンパク質の部分は、溶液中で遊離しているか、非生物基板または生物基板に固定され（例えば細胞表面上に保持される）、あるいは細胞内に位置しうる。例えば1つの方法では、生存可能または固定された原核宿主細胞で、その細胞表面にペプチドを発現し位置させている組換え核酸で安定的に形質転換される細胞を、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することが可能である。スクリーニングする分子または化合物の種類に応じ、アッセイを用いて、アゴニスト、アンタゴニスト、抗体、DNA分子、小分子薬、イムノグロブリン、インヒビター、擬態物質、ペプチド、ペプチド核酸、タンパク質、およびRNA分子、または任意の他のリガンドであって、該タンパク質に特異結合するものを同定できる。

10

20

【0128】

一態様で本発明は米国特許第5,876,946号に記載されるような、極めて小さなアッセイ容量と極めて少量の試験化合物を用いる或る高処理スクリーニング法を企図し、この特許は参照により開示に含まれる。この方法は、特異結合によって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。別の様態において本発明は、タンパク質に特異結合可能な中和抗体が該タンパク質に結合可能な試験化合物と競合する、競合薬スクリーニングアッセイの使用も企図する。スクリーニングによって同定した分子または化合物を用いて、哺乳類モデル系で、その毒性または治療可能性を評価できる。

【0129】

医薬組成物

医薬組成物を処方し、そうした治療を要する被検者に投与して、治療効果を果たすことができる。そうした組成物には、即効性タンパク質、アゴニスト、アンタゴニスト、二重特異性分子、小分子薬、イムノグロブリン、インヒビター、擬態物質、多重特異性分子、ペプチド、ペプチド核酸、医薬品、タンパク質、およびRNA分子を含む。組成物の製造には、従来手段たとえば混合、溶解、顆粒化、糖衣形成、微粒化（levigating）、乳化、カプセル化、封入（entrapping）、または凍結乾燥を行う。該組成物は酸たとえば塩酸、硫酸、酢酸、乳酸、酒石酸、リンゴ酸、琥珀酸で形成した塩として、または凍結乾燥粉末として提供され、これは無菌バッファー例えば食塩水、デキストロース、または水と併用されうる。これらの組成物には、活性化合物のプロセッシングを促進する補助剤または賦形剤を含みうる。

30

40

【0130】

補助剤および賦形剤としては、コーティング、充填剤、または結合剤があり、これには糖たとえばラクトース、スクロース、マンニトール、グリセロール、またはソルビトールや、トウモロコシ、小麦、米、またはジャガイモの澱粉や、タンパク質たとえばアルブミン、ゼラチン、コラーゲンや、ヒドロキシプロピルメチルセルロース、メチルセルロース、またはカルボキシメチルセルロースナトリウムの形のセルロースや、アラビアゴムとトラガカントなどゴムや、潤滑剤たとえばステアリン酸マグネシウムまたは滑石や、崩壊剤または可溶化剤たとえば寒天、アルギン酸、アルギン酸ナトリウムまたは架橋ポリビニルピロリドンや、安定剤たとえばカルボポール（carbopol）ゲル、ポリエチレングリコール、二酸化チタンや、染料または色素があり、これらは産物を同定し活性化合物の含量、または

50

用量を特徴付けるため添加する。

【0131】

これらの組成物は、任意の数の経路によって投与でき、経路には、経口、静脈内、筋肉内、動脈内、髄内、髄腔内、心室内、経皮、皮下、腹腔内、鼻腔内、腸内、局所、舌下または直腸がある。

【0132】

投与の経路と用量が剤形を決めることとなる。例えば経口投与には錠剤、丸剤、糖衣、カプセル、液体、ゲル、シロップ、スラリー、または懸濁液を用い、非経口投与は水性で生理適合バッファー例えばハックス液、リングル液、または緩衝生理食塩水で処方されうる。注射用懸濁液は、水性で粘稠添加物たとえばカルボキシメチルセルロースナトリウムまたはデキストランを含有して粘稠度を増した液や、油性で親油性溶媒たとえば胡麻油または合成脂肪酸エステルたとえばオレイン酸エチルまたはトリグリセリド、あるいはリポソームを有する液でありうる。当該分野で公知の浸透剤は、局所または鼻内投与に用いる。

10

【0133】

毒性および治療有効性

治療有効量とは、症候または病状を寛解させる活性分量を言う。任意の化合物に対して、細胞培養アッセイに正常および腫瘍性の細胞を用いて、或いは、動物モデルで、治療有効量を推定できる。治療有効性、毒性、濃度範囲、および投与経路を決定するには、実験動物を用いた標準の製薬手順を用い得る。

【0134】

治療指数は治療効果と毒性効果の用量比、LD50（集団の50%の致死量）/ED50（集団の50%の治療有効量）であり、大きな治療指数が好ましい。投与量は循環濃度の範囲内で、ED50を含み毒性はわずかまたは無く、組成、送達法、患者の感受性、および投与の経路によって変動する。正確な投与量は、治療が必要な被検者に関する要素を考慮して、実務医が決定することになる。

20

【0135】

用法および用量は調整され、目的は治療効果を維持する活性成分の提供である。調整の要素としては、病態の重症度、患者の全身健康状態、患者の年齢、体重及び性別（ジェンダー）、食事、投与の時間及び頻度、併用薬、反応感受性及び治療に対する許容度/応答等がある。長時間作用性の医薬組成物は、特定組成物の半減期及びクリアランス速度によって3~4日毎に1度、毎週、或いは隔週の間隔で投与し得る。

30

【0136】

通常投与量は、投与経路によるが0.1µgから最大で総用量約1gまで変動しうる。特定組成物の用量は、別の媒介物、薬物、またはホルモンと併用して患者に投与する際は、低くなりうる。特定の用量決定および送達法に関する指針は、医薬文献に示される。製剤および投与の技術の詳細は、Remington's Pharmaceutical Sciences (Mack Publishing, Easton PA) の最新版にも見られる。

【0137】

モデル系

動物モデルは生物検定として用い得る。そこではヒトに類似の表現型応答を示し、また暴露条件が人体暴露と関連する。哺乳類は最も一般的なモデルであり、大抵の感染因子、癌、薬物、及び毒性の試験は齧歯類たとえばラットやマウスで行うが、理由は、低コスト、可用性、寿命、妊娠期間、子孫の数、及び豊富な参照文献である。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的結果の研究、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）は、その遺伝子により発現するタンパク質の便利な源泉となり得る。

40

【0138】

毒性研究

毒性研究は、生物系への薬剤影響の研究である。大多数の毒性研究は、ラットまたはマウ

50

スで実施される。ラットまたはマウスでの、生理、行動、恒常性プロセス、および致死性での定性または定量的変化の観察を用いて毒性プロファイルが作成され、該薬剤に曝露後のヒトの健康への影響がアセスされる。

#### 【0139】

遺伝毒性研究により、内因性、自然発生および誘導された遺伝子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は機能的異常の頻度が増大する薬剤を同定することが可能である。マウスまたはラットは、これらの試験で最も頻繁に使用される。なぜなら生殖周期が短いので統計上の必要にかなう多数の生物の繁殖を可能にするからである。

10

#### 【0140】

急性毒性試験は、薬剤の症候学または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。すなわち、1) 初回の用量範囲探索試験、2) 有効量の範囲を狭める試験、3) 用量反応曲線を確定するための最終試験、である。

#### 【0141】

亜慢性毒性試験は、薬剤の反復投与に基づく。ラットとイヌはこれらの研究で一般的に用いられ、異なる科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、高用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。

20

#### 【0142】

慢性毒性試験の期間は1年以上であり、これを用いて、長期投与が毒性、催奇形性、または発癌を誘発するかどうか試験する。ラットで研究が実行される時は、最少で3つの試験群に加え1つの対照群が用いられる。実験の開始時および定期的に動物が試験され、モニターされる。

#### 【0143】

##### 遺伝子組換え動物モデル

当該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、ヒト疾患のモデル化、または治療薬や毒剤の試験のために用いられ得る(米国特許第5,175,383号、第5,767,337号等を参照)。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、試験的薬物治療への取り組み前中後に遺伝子組換え動物の、表現型、組織特異的mRNA発現、または血清と組織タンパク質レベルの分析によってモニターされる。

30

#### 【0144】

##### 胚性幹細胞

齧歯類の胚から単離した胚性幹細胞(ES)は、胚組織を形成する能力を保持する。ES細胞はキャリアの胚内に置かれる時、正常な発育を再開して、生きて生まれた動物の組織に寄与する。ES細胞は実験的ノックアウトおよびノックイン齧歯類系の生成に用いられる好適な細胞である。マウスES細胞たとえばマウス129/SvJ細胞株は早期マウス胚に由来し、当該分野で公知の培養条件で成長する。トランスジェニック系の産生に用いるベクターは疾患遺伝子候補と標識遺伝子を有し、後者は導入した疾患遺伝子の存在を同定するよう働く。ベクターは当該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。これらの胚盤胞を偽妊娠メスに外科的移植し、得られるキメラ子孫の遺伝子型を決定し、これらを交配させてヘテロ接合性系またはホモ接合性系を作製する。

40

#### 【0145】

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

50

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。

【0146】

#### ノックアウト分析

遺伝子ノックアウト分析では、遺伝子の或る領域を酵素的に修飾し、非哺乳類遺伝子たとえば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) などを含める。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、内因性遺伝子の転写と翻訳を攪乱する。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的コピーを欠くホモ接合性近交系を得るために交雑される。一例では、哺乳動物遺伝子はヒト遺伝子である。

10

【0147】

#### 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を試験し、医薬品を用いて処理し、類似のヒト病態の治療に関する情報を得る。これらの方法は、幾つかのヒト疾患をモデル化するために用いられてきた。

20

【0148】

#### 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、ヒト以外の霊長類での治療薬の影響評価に最も重要である。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。またサルの反応は類似条件下の人体暴露と関連する。このような調査では、カニクイザルとアカゲザル(それぞれ *Macaca fascicularis* と *Macaca mulatta*)、コモンマーモセット (*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の群体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。薬物嗜癖などの行動測定を利用する研究では、NHPが第一選択の実験動物である。またNHPと個々のヒトは多くの薬物と毒素に差次的感受性を示し、或る範囲の表現型、これら物質の「高代謝群」から「低代謝群」に分類できる。

30

【0149】

別の実施態様では、該タンパク質をコードするcDNA群を、今後開発される分子生物学技術であり、現在知られているcDNA群の性質(限定はされないが、例えばトリプレット暗号、特異的な塩基対相互作用等)に依存する任意の新技术に用い得る。

【実施例】

【0150】

#### 1 cDNAライブラリの作製

COLNTUT03ライブラリは62歳の白人男性のS状結腸切除および永久的人工肛門形成時にS状結腸から得た結腸腫瘍組織から単離したRNAを用いて作製した。病理はグレード2の腺癌を示し、これは筋層に浸潤していた。

40

【0151】

POLYTRONホモジナイザー (Brinkmann Instruments, Westbury NJ) を使用して、凍結組織をホモジナイズしグアニジニウムイソチオシアネート溶液に溶解した。可溶化液を、L8-70M超遠心機 (Beckman Coulter, Fullerton CA) でSW28ローターを使用して5.7Mの塩化セシウムクッション上で18時間、25,000 rpm、周囲温度で遠心分離した。RNA抽出には酸性フェノール (pH 4.7) を用い、沈殿には0.3 M酢酸ナトリウムと2.5倍量のエタノールを用い、再懸濁をRNaseフリー水で行い、DNase処理を37 °Cで行った。酸性フェノール (pH 4.7) 抽出と酢酸ナトリウムおよびエタノールでの沈殿を繰り返した。mRNAをOLIGOTEXキット

50

(Qiagen, Chatsworth CA) で単離して、cDNAライブラリの作製のために用いた。

【0152】

mRNAの処理はSUPERSCRIP Tプラスミドシステム (Life Technologies) の推奨プロトコルに従い、これはmRNAのポリ(A)尾部で第1鎖cDNA合成を始めるよう設計されたNotIプライマー-アダプタを有する。二本鎖cDNAを平滑末端化し、EcoRIアダプタに連結し、NotI (New England Biolabs, Beverly MA) で消化した。cDNAをSEPHAROSE CL4Bカラム (APB) 上で分画して、そのうち400bpを超えるcDNAはpINCY プラスミド (Incyte Genomics) に連結させた。プラスミドpINCYを次に、DH5 コンピテント細胞群 (Life Technologies) に形質転換した。

【0153】

10

2 cDNAの単離、調製、およびシーケンシング

プラスミドの、宿主細胞からの回収は、UNIZAPベクターシステム (Stratagene) を用いた *in vivo* 切除によって、あるいは細胞溶解によって行った。プラスミドを精製する方法は、MagicまたはWIZARD Minipreps DNA精製システム (Promega)、AGTC Miniprep精製キット (Edge Biosystems, Gaithersburg MD)、Qiagen社のQIAWELL 8 Plasmid、QIAWELL 8 Plus Plasmid、QIAWELL 8 Ultra Plasmid 精製システム、REAL PREP 96プラスミド精製キットの中から少なくとも1つを用いた。プラスミドは、沈殿させた後、0.1mlの蒸留水に再懸濁して、凍結乾燥して或いは凍結乾燥しないで4 で保管した。

【0154】

別法では高処理フォーマットで直接結合PCRを用い、宿主細胞溶解物からプラスミドDNAを増幅した (Rao (1994) Anal Biochem 216:1-14)。宿主細胞の溶解および熱サイクリング過程は、単一反応混合液中で行った。サンプルをプロセスして384穴プレート内で保管し、増幅したプラスミドDNAの濃度をPICOGREEN色素 (Molecular Probes, Eugene OR) およびFLUOROSKAN II蛍光スキャナ (Labsystems Oy, Helsinki, Finland) を用いて蛍光分析的に定量した。

20

【0155】

シーケンス反応は、標準的方法あるいは高処理装置、例えばCATALYST 800 (ABI) サーマルサイクラーまたはDNA ENGINEサーマルサイクラー (MJ Research) を、HYDRAマイクロディスペンサー (Robbins Scientific) またはMICROLAB 2200 (Hamilton) 液体転移システムと併用して処理した。cDNAのシーケンス反応は、APB社が提供する試薬、またはシーケンシングキット、例えばPRISM BIGDYE Terminator cycle sequencing ready reaction kit (ABI) の試薬を用いて準備した。cDNAシーケンス反応の電気泳動分離と、標識したポリヌクレオチドの検出とは、MEGABACE 1000 DNAシーケンシングシステム (APB) が、PRISM 373または377シーケンシングシステム (ABI) を、標準プロトコルとベースコーリングソフトウェアと共に用いた。cDNA配列内のリーディングフレームの同定には、標準の方法を用いた (Ausubel前出、ユニット7.7)。

30

【0156】

3 cDNAの伸長

cDNAクローンとオリゴヌクレオチドプライマー対を用いて、cDNAを伸長した。一方のプライマーは既知断片の5'伸長を開始するべく合成し、他方のプライマーは既知断片の3'伸長を開始するべく合成した。開始プライマーの設計にはプライマー分析ソフトウェアを用い、長さが約22~30ヌクレオチド、GC含量が約50%以上、温度約68 ~ 約72 で標的配列にアニーリングするようにした。ヘアピン構造及びプライマー-プライマー二量体が生じないようにヌクレオチドを伸長した。

40

【0157】

選択したcDNAライブラリ群を鋳型として用い、配列を伸長した。2回以上の伸長を行った場合には、付加的プライマーあるいはプライマーのネステッドセットを設計した。好適なライブラリ群のサイズを選択してより大きなcDNA群を含むようにし、遺伝子群の5'または上流領域を持つ配列を多く有するようにランダムプライムした。ゲノムライブラリ群は、5'プロモータ結合領域に伸長する調節エレメントを得るために用い得る。

50

## 【0158】

高忠実度の増幅を、米国特許第5,932,451号で開示されたような方法を利用したPCR法によって得た。PCR法は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を有し、また、200 nmolの各プライマーを有する。また、 $Mg^{2+}$ と $(NH_4)_2SO_4$ とメルカプトエタノールとを含有する反応バッファート、Taq DNAポリメラーゼ (APB) と、ELONGASE酵素 (Invitrogen) と、Pfu DNAポリメラーゼ (Stratagene) とを含有する。プライマー対であるPCI AとPCI B (Incyte Genomics) に対し、以下のパラメータで増幅した。各サイクルのパラメータは、1: 94 で3分; 2: 94 で15秒; 3: 60 で1分; 4: 68 で2分; 5: 2, 3, および4を反復20回; 6: 68 で5分; および7: 4 で保存である。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。1: 94 で3分; 2: 94 で15秒; 3: 57 で1分; 4: 68 で2分; 5: 2, 3, および4を反復20回; 6: 68 で5分; および7: 4 で保存。

10

## 【0159】

各ウェルのDNA濃度は、100  $\mu$ lのPICOGREEN定量試薬 (0.25% 試薬を1  $\times$  TE中 (v/v); Molecular Probes) および0.5  $\mu$ lの希釈していないPCR産物を不透明な蛍光光度計プレート (Corning Life Sciences, Acton MA) の各ウェルに分配し、DNAが試薬と結合できるようにして判定した。サンプルの蛍光を計測してDNAの濃度を定量すべく、プレートをFluoroskan II (Labsystems Oy, Helsinki Finland) でスキャンした。反応混合物のアリコット5~10  $\mu$ lを1%アガロースミニゲル上の電気泳動によって解析し、どの反応が配列の伸長に成功したかを決定した。

20

## 【0160】

伸長したクローンは、脱塩及び濃縮して384穴プレートに移し、CviJIコレラウイルスエンドヌクレアーゼ (Molecular Biology Research, Madison WI) を用いて消化し、pUC 18ベクター (APB) への再連結反応前に音波処理またはせん断した。ショットガン・シーケンシングのために、消化したヌクレオチド配列を低濃度 (0.6~0.8%) のアガロースゲル上で分離し、断片を切除し、寒天をAGARACE酵素 (Promega) で消化した。伸長したクローンをT4DNAリガーゼ (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オーバーハングを満たし、大腸菌コンピテント細胞に形質移入した。形質転換した細胞群の選択を抗生物質含有培地で行い、それぞれのコロニーを採取しLB/2 $\times$ カルベニシリン液体培地の384穴プレートに37で一晚培養した。

30

## 【0161】

細胞を溶解し、プライマー、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を増幅した。1: 94 で3分; 2: 94 で15秒; 3: 60 で1分; 4: 72 で2分; 5: 2, 3, および4を反復29回; 6: 72 で5分; および7: 4 で保存。上記したようにPICOGREEN定量試薬 (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エネルギー移動シーケンシングプライマー、およびDYENAMIC DIRECTサイクル シーケンシングキット (APB) またはPRISM BIGDYEターミネーターサイクル シーケンシングキット (ABI) を用いてシーケンシングした。

40

## 【0162】

## 4 cDNAクローンとその推定タンパク質の相同性検索

配列表のcDNA、またはその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。過去に同定され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BLASTかBLAST2を用いて検索してアラインメントを生成し、またどの配列が厳密に一致するか、または相同体かを決定した。アラインメントは原核生物 (細菌) または真核生物 (動物、真菌あるいは植物) 起源の配列へ行った。別法として、SmithとSmith (1992, Protein Engineering 5:35-51) で説明されてい

50

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を扱うことが可能であった。本願に開示す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12%に過ぎない（A、C、GまたはTでなくNが記録されている）。

#### 【0163】

KarlinとAltschul（1993; Proc Natl Acad Sci 90:5873-5877）で詳述されているように、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、ヌクレオチドで $10^{-25}$ 、ペプチドで $10^{-14}$ の閾値を満たす時のみ報告させた。相同性の評価はまた、以下のように計算した積スコアで行った。BLASTでの%ヌクレオチドまたはアミノ酸同一性[問合せ配列と参照配列の間]に%最大可能BLASTスコア[問合せ配列と参照配列の長さに基づく]を乗じ、次に100で除す。ラボで用いるハイブリッド形成手順に比し、厳密な一致のストリンジェンシーは、下限である約40（不要な塩基による1-2%の誤差あり）から、100%一致の約70までに設定した。

10

#### 【0164】

BLASTソフトウェア（NCBI, Bethesda MD）には様々な配列分析プログラムが含まれ、これにはヌクレオチド配列をアラインメントするのに用いる「blastn」と、ヌクレオチドまたはアミノ酸配列の直接ペアワイズ比較に用いるBLAST2がある。BLASTプログラムは、一般的には、ギャップ（gap）などのパラメータをデフォルト設定にセットして用いる。例えば以下のとおりである。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; およびFilter: on。同一性は配列全体の長さについて測定する。Brenner（前出）はBLASTが配列同一性によって構造相同体を同定する能力を分析し、30%の同一性が150残基以上の配列アラインメントでの信頼できる閾値であり、40%が70残基以上のアラインメントの閾値であると発見した。

20

#### 【0165】

本願のcDNA群を、LIFESEQ GOLDデータベース（Incyte Genomics）で見られるアセンブリされたコンセンサス配列あるいは鋳型と比較した。cDNA、伸長、全長およびショットガン・シーケンシングプロジェクトのコンポーネント配列をPHRED分析にかけ、質スコアを割り当てた。質スコアが許容できるすべての配列を多様な予備処理経路と編集経路にかけ、低品質の3'末端、ベクターとリンカー配列、ポリAテイル、Alu反復、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した。編集した配列は少なくとも長さが50 bpとし、情報に乏しい配列と、ジヌクレオチド反復、Alu反復などの反復エレメントを「N」に置換あるいはマスクした。

30

#### 【0166】

編集した配列はアセンブリ処理にかけられ、配列が遺伝子ビン（gene bins）に割り当てられた。各配列は1つのビンにのみ所属でき、各ビンの配列群をアセンブリして鋳型を製作した。BLASTとCROSSMATCHを用いて、新たにシーケンスしたコンポーネント群を既存のビンに加えた。ビンに加えるには、前記のコンポーネント配列は150以上のBLAST質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することとした。PHRAPを用いて、各ビンの配列をアセンブリした。DEEP PHRAPを用いて、オーバーラップコンポーネント配列を幾つか有するビンをアセンブリした。コンポーネント配列の数と配向に基づいて、各鋳型の配向を決定した。

40

#### 【0167】

ビ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ビンを結合し再アセンブリした。95%未満の局所同一性である鋳型を有するビンは分割した。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、選択的スプライシングされたエキソン、スプライス接合部や、組織の各タイプや病態全体での選択的スプライシングされた遺伝子の差次的発現などの、存在の確率を決定した。アセンブリ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。厳密な一致の定義は95%の局所同一性が200塩基対に及ぶものから100%の局所同一性が100塩基対に及ぶものとし、相同性一致はE値（すなわち確率スコア）が

50

$1 \times 10^{-8}$ とした。鋳型はまた、フレームシフトFASTxをGENPEPTに対してかけ、相同性一致の定義はE値が  $1 \times 10^{-8}$ とした。鋳型の分析とアセンブリについては、1999年3月25日に出願の米国特許出願第09/276,534号に記載がある。

#### 【0168】

アセンブリ後、鋳型をBLAST, motif, 及び他の機能解析にかけ、タンパク質の各ヒエラルキーに分類したが、其の方法は、共に1997年3月6日に米国特許出願第08/812,290号および08/811,758号; 1997年10月9日に米国特許出願第08/947,845号; および1998年3月4日に米国特許出願第09/034,807号に記載の方法を用いた。次に鋳型の分析のため、各鋳型の翻訳を全部で3の順行リーディングフレームで行い、各翻訳の検索を隠れマルコフモデル系蛋白ファミリー群およびドメイン群のPFAMデータベースに対しHMMERソフトウェアパッケージ (Washington University School of Medicine, St. Louis MO) を用いて行った。MACDNASIS PROソフトウェア (Hitachi Software Engineering) とLASERGE NEソフトウェア (DNASTAR) を用いてcDNAを更に分析した。そして公共のデータベース、例えばGenBankのげっ歯類、哺乳類、脊椎動物、原核生物、真核生物のデータベースと、SwissProt、BLOCKS、PRINTS、PFAMそしてPrositeに問い合わせた。

10

#### 【0169】

### 5 ノーザン分析、転写イメージング、および「Guilt-By-Association」法

#### ノーザン分析

ノーザン分析は遺伝子の転写物の存在の検出に用いる実験技術であり、特定の細胞種または組織からのRNAが結合している膜への、標識されたヌクレオチド配列のハイブリダイゼーションに参与する。本技術は下記の実施例7とAusubel (前出、ユニット4.1-4.9) に記載されている。

20

#### 【0170】

類似の、BLASTを応用したコンピュータ技術を用いて、GenBankまたはLIFESEQデータベース (Incyte Genomics) などのヌクレオチドデータベースにおいて、同一または関連分子を検索する。ノーザン分析は、幾つかの膜系ハイブリダイゼーションよりも高速である。更に、任意の特定の一致を厳密なあるいは相動的なものとして分類するか否かを決定するため、コンピュータ検索の感度を修正できる。検索の基本は積スコアであり、上記の実施例4に記載した。

#### 【0171】

ノーザン分析の結果の報告は、TMDCをコードする転写物がそこで生じるライブラリ群のリストとして成される。存在量とパーセント存在度も報告される。存在量は直接に、特定転写物が或るcDNAライブラリ内で示された回数を反映し、パーセント存在度は、存在量を、該cDNAライブラリで試験した配列の総数で除した数である。

30

#### 【0172】

#### 転写イメージング

転写イメージ (transcript image) 作成にはLIFESEQ GOLDデータベース (Incyte Genomics) を用いた。本プロセスにより全cDNAライブラリで発現されるポリヌクレオチド群の相対的存在度のアセスメントができ、これは米国特許第5,840,484号に記載され、参照により開示に含まれる。LIFESEQデータベースの全ての配列とcDNAライブラリは、系、臓器/組織および細胞タイプで分類される。カテゴリは、心血管系、結合組織、消化器系、胎児構造、内分泌系、外分泌腺、女性生殖器、男性生殖器、生殖細胞、血液および免疫系、肝、筋骨格系、神経系、脾臓、呼吸器系、感覚器、皮膚、顎口腔系、非分類性/混合性、および尿路がある。転写イメージングの判定基準は、カテゴリ、ライブラリ当りのcDNA数、ライブラリの説明、疾患適応、サンプルの臨床的意義等から選ぶ。

40

#### 【0173】

各カテゴリにつき、そこで該配列が発現されたライブラリの数がかウントされ、そのカテゴリのライブラリの総数に対して示された。各ライブラリにつき、cDNA数がかウントされ、該ライブラリ内の総cDNA数に対して示された。転写イメージによっては、全ての濃縮した、ノーマライズした、またはサブトラクトしたライブラリ (高いコピー数の配列を持つ

50

)を除去した後にプロセッシングでき、また全ての混合組織またはプールした組織(複数の組織タイプまたは複数被検者の組織を持つという意味で非特異的と見なされる)を分析から除外できる。処理済および無処理の細胞株および/または胎児組織のデータも、臨床的意義が強調される際に除外できる。逆に胎児組織を強調でき、これは遺伝性障害の解明または、特定の成体幹細胞または胚性幹細胞を組織または器官(例えば心臓、腎臓、神経または膵臓)へ分化させる際に行い、これを補助するため臨床サンプルを分析から除外することとなる。転写イメージはまた、他の方法たとえばハイブリッド形成、「guilt-by-association」法、およびアレイ技術からのデータの補助に用い得る。

【0174】

#### 「Guilt-By-Association」法

GBA法は、或る特定の疾患プロセス、細胞内区画、細胞タイプ、組織タイプ、または種に関連する複数のcDNAライブラリで発現されるcDNA群を同定する。未知の機能を持つcDNAの発現パターンが、十分に考証された機能を持つ遺伝子群の発現パターンと比較され、指定した共発現確率閾値に合うか判定される。この比較を通じて、高度に有意な共発現確率を既知の遺伝子群に対して持つcDNAのサブセットを同定する。

【0175】

該cDNAの起源はヒトcDNAライブラリ群であり、これは任意の細胞または細胞株、組織、または器官に由来し、種々の配列タイプから選択でき配列としては、限定はされないが、発現配列タグ(ESTs)、アセンブリしたポリヌクレオチド、全長遺伝子コード領域、プロモータ、イントロン、エンハンサ、5'非翻訳領域、および3'非翻訳領域がある。統計的に有意な分析結果を得るために、少なくとも5つのcDNAライブラリで該cDNAを発現する必要がある。配列を分析するcDNAライブラリの数の範囲は、少ないもので500から10,000以上でありうる。

【0176】

統計的に有意な共発現パターンを示すcDNAを同定する方法は以下のとおりである。最初に、或るcDNAライブラリ内の或る遺伝子の存在あるいは不存在を定義する。或る遺伝子に対応する少なくとも1つの、その配列の断片が、或るライブラリから採取した或るサンプルに検出される時、該遺伝子はそのライブラリに存在する。そして対応する断片がサンプル中で検出されない時、該遺伝子はそのライブラリに存在しない。

【0177】

2番目に、確率法を用いて共発現の有意性を評価して、共発現の偶然による確率を測定する。確率法はフィッシャーの正確確率検定、カイ2乗検定またはカッパ検定であり得る。これらの検定と適用例は当該分野で公知であり、標準的な統計学のテキストで見出すことができる(Agresti(1990) *Categorical Data Analysis*, John Wiley & Sons, New York NY; Rice(1988) *Mathematical Statistics and Data Analysis*, Duxbury Press, Pacific Grove CA)。ボンフェローニ補正(前出のRice、384ページ)も、1つの遺伝子の統計結果を多数の他の遺伝子に対して補正するために、確率法の1つと併用して適用できる。或る好適な実施態様では、偶然による確率をフィッシャーの正確確率検定により測定する。また偶然による確率の閾値を好ましくは0.001未満に設定する。

【0178】

2つの遺伝子の共発現の確率を推計するこの方法の前提は、ライブラリ群が独立しており同様にサンプルされることである。しかし実際の状況では、選択したcDNAライブラリは完全には独立していない。理由は以下である。1)複数ライブラリを単一の対象または組織から得ることができる。そして2)様々な数のcDNAを、通常の範囲は5000~10000で、各ライブラリからシーケンスできるからである。更に、フィッシャーの正確共発現確率を、各遺伝子に対し、少なくとも5つのライブラリで出現する他のすべての遺伝子ごとに計算するので、複数の統計試験のためのボンフェローニ検定を用いた(Walker 他(1999; *Genome Res* 9:1198-203; 特に参照により開示に含まれる)。

【0179】

## 6 染色体マッピング

10

20

30

40

50

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、配列表に存在する任意のcDNAが過去にマッピングされたかを判定した。TMDCをコードするcDNAの何れかの断片がマップされていれば、全ての関連する調節配列とコード配列を同じ場所に割り当てる。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単位 (ヒトDNAの1メガベースにほぼ等しい) での或る間隔の地図上の位置を、染色体短腕 (p-arm) の末端に対して測定する。

【0180】

## 7 ハイブリッド形成および増幅の技術ならびに分析

### 組織サンプルの調製

正常および癌組織のサンプルは、下表のドナー識別番号で記載する。列1はドナーID、列2はドナーの年齢/性別、列3は障害の説明、列4は腫瘍の分類、列5は出所を示す。

ドナー	年齢 / 組織および説明	病期	出所
3579	55/M 結腸、高分化腺癌	Dukes' C; TMN T2N1	HCI
3580	38/M 結腸、低分化、転移性腺癌	T3N1MX	HCI
3581	U/M 直腸、腫瘍	NA	HCI
3582	78/M 結腸、中分化腺癌	TMN T4N2MX	HCI
3583	58/M 結腸、腺管絨毛腺腫 (過形成ポリープ)	NA	HCI
3647	83/U 結腸、浸潤性中分化腺癌 (管状腺腫)	TMN T3N1MX	HCI
3649	86/U 結腸、浸潤性高分化腺癌	NA	HCI
3479	68/M 結腸、腺癌	NA	HCI
3839	59/M 結腸腫瘍	U	HCI
4614	67/U 結腸、中分化腺癌	Dukes' B; TMN T3N0	HCI

\*略語: CA=癌、U=不明、NA=入手不可

【0181】

図3で、ノーマライズした第1鎖合成、cDNA標本は正常なヒトの心臓、全脳、肺、肝臓、骨格筋、腎臓、膵臓、脾臓、胸腺、前立腺、卵巣、小腸、末梢血白血球、および大腸の組織からの標本であり、Clontechから得た。更なるcDNA標本はヒト成人の正常な甲状腺、下垂体、および副腎組織の標本であり、Clinomics Bioscience (Pittsfield MA) から得た。

【0182】

図5に示す大腸直腸腺癌細胞株群はATCCから得た株であり、供給者の規格に従って培養した。細胞株は、LS123、LS174T、HCT116、CaCo2、HT29、SW480、Colo205、T84、およびSW620である。

【0183】

### 基板上でのcDNAの固定

下記の方法の1つによってcDNAを基板に付ける。ゲル電気泳動法によりcDNAの混合物を分画し、キャピラリー転写 (capillary transfer) によりナイロン膜に転写する。別法として、cDNAを個々にベクターに連結し、細菌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を基板に配置する。第一法では、個々のクローンを有する細菌細胞をロボットで採取し、ナイロン膜に配置する。膜を選択剤 (用いる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、コロニー側を上にして、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、継続的に膜を移す。次に膜をSTRATALINKER UVクロスリンク機 (Stratagene) でUV照射する。

【0184】

第2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群を用いて、cDNAを30サイクルのPCRで細菌ベクターから増幅する。PCR増幅は1~2 ngの開始濃度の核酸を5 µgより大きい最終量まで増加する。長さ約400 bpから約5000 bpまで増幅した核酸を、SEPHACRYL-400ビーズ (APB) を用いて精製する。手であるいはドット/スロットブロット法マニホールドと吸引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、およびUV照射により固定する。米国特許第5,807,522号に記載の方法を用いて、精製核酸をロボットで配置し、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

10

20

30

40

50

鏡スライドガラス (Corning Life Sciences) を洗浄するため0.1%のSDSおよびアセトン中で超音波をかけ、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し、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を準備する。スライドは処理の間および処理後に十分に蒸留水で洗浄する。核酸をスライド上に置き、次にSTRATALINKER UVクロスリンク機 (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。アレイを次に室温において0.2% SDSで洗浄し、蒸留水中で3回すすぐ。非特異結合部位をブロックするためリン酸緩衝食塩水 (PBS; Tropix, Bedford MA) 中に0.2%カゼイン、60 °Cで30分間アレイをインキュベートした後、前記のように0.2% SDS中で洗浄し蒸留水中ですすぐ。

10

【0185】

#### 膜ハイブリダイゼーションのプロープ調製

配列表のcDNA群に由来するハイブリダイゼーション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。プロープを調製するため、cDNAを45 µl TEバッファー中40~50 ngの濃度に希釈し、5分間100 °Cに加熱して変性させ、短時間遠心分離する。次に変性したcDNAをREDIPRIMEチューブ (APB) に加え、青色が均一に分布するまで穏やかに混合して、次に短時間遠心分離する。5 µlの<sup>32</sup>P]dCTPを管に加え、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停止する。そしてPROBEQUANT G-50マイクロカラム (APB) を用いてプロープを、組み込まれていないヌクレオチドから精製する。精製したプロープを5分間100 °Cに加熱し、2分間氷上で急冷して、以下に記載した膜系ハイブリダイゼーションで使用する。

20

【0186】

#### QPCR用のプロープ調製

QPCR用プロープをABI社のプロトコルに従い調製した。

【0187】

#### ポリマーコートされたスライドでのハイブリダイゼーションのプロープ調製

次の方法を使用して、図3に示すマイクロアレイ分析のプロープの調製を行った。サンプル群から単離したmRNAに由来するハイブリダイゼーションプロープを使用して、アレイベースのハイブリダイゼーションで配列表のcDNAをスクリーニングする。GEMbrightキット (Incyte Genomics) を用い、9 µl TE バッファーで200 ngの濃度にmRNAを希釈し、5 µl 5x バッファー、1 µl 0.1 MのDTT、3 µl Cy3またはCy5の標識ミックス、1 µl RNアーゼ阻害剤、1 µl逆転写酵素、5 µl 1x 酵母対照mRNAを加えてプロープを調製する。非コード酵母ゲノムDNAから *in vitro* 転写により酵母対照mRNA群を合成する (W. Lei, 未発表)。定量用対照として、0.002 ng、0.02 ng、0.2 ngおよび2 ngでの対照mRNAの1セットを、対サンプルmRNA比がそれぞれ1:100,000、1:10,000、1:1000、1:100 (w/w) で逆転写反応混液に希釈する。mRNA差次的発現パターンを調べるため、対照mRNAの第2セットを1:3、3:1、1:10、10:1、1:25および25:1 (w/w) の比で逆転写反応混液に希釈する。反応混液を混合し、37 °Cで2時間インキュベートする。次に反応混液を20分間、85 °Cでインキュベートする。そして2つの連続するCHROMA SPIN+TE 30カラム (Clontech, Palo Alto CA) を用いてプロープを精製する。精製したプロープのエタノール沈殿を、プロープをDEPC処理水中で90 µlに希釈し、2 µl 1mg/mlグリコーゲン、60 µl 5 M酢酸ナトリウム、300 µl 100% エタノールを加えて行う。プロープを20分間、20,800xgで遠心分離する。そしてペレットを12 µlの再懸濁バッファーで再懸濁し、5分間65 °Cに加熱してから、十分に混合する。プロープを加熱し、前記のように混合して、氷上に保管する。以下に述べるように、プロープを高密度アレイベースのハイブリダイゼーションにおいて使用する。

30

40

【0188】

#### in situハイブリダイゼーション

*in situ*ハイブリッド形成を用いて、切片組織での膜貫通タンパク質の発現を判定した。ジゴキシゲニンプロトコルで、新鮮な凍結切片 (10ミクロン厚) をフリーザーから移し、直ちに浸漬を4%パラホルムアルデヒドで10分行い、PBSですすぎ、アセチル化を0.1 M TE

50

A, pH 8.0に0.25% (v/v) 無水酢酸で行う。組織を5×SSCに平衡させた後、プレハイブリダイゼーションをハイブリッド形成バッファー (50%ホルムアミド、5×SSC、1×デンハート液、10%デキストラン硫酸、および1 mg/mlニシン精子DNA)で行う。

#### 【0189】

ジゴキシゲニン標識したTMDC特異的RNAプローブである、SEQ ID NO:2のcDNAから選択したセンスおよびアンチセンスのヌクレオチドの産生は以下のようにした。1) pINCYプラスミドであってSEQ ID NO:2のほぼヌクレオチド1068からほぼヌクレオチド2324に及びSEQ ID NO:2の或る断片 (1519 bp) を有するプラスミドの直線化をEcoRI (アンチセンス) またはNotI (センスプローブ)で行い、2) *in vitro* 転写にはT7 (アンチセンス) またはSP6 (センス) RNAポリメラーゼを用い、3) 加水分解して平均の長さを350 bpにした。約500 ng/mlのRNAプローブを用いて一晩のハイブリッド形成を、65 °C、ハイブリッド形成バッファーで行った。ハイブリッド形成後、切片のすすぎを30分間2×SSC、室温で、1時間2×SSC、65 °Cで、また1時間0.1×SSC、65 °Cで行った。切片をPBSで平衡化し、プロッキングを30分10% DIGキットブロッカー (Roche Molecular Biochemicals, Indianapolis IN) を含むPBSで行った後、インキュベートを一晩4 °C、1:500抗-DIG-AP中で行った。翌日、切片をPBSですすぎ、平衡化を検出バッファー (0.1 M Tris, 0.1 M NaCl, 50 mM MgCl<sub>2</sub>, pH 9.5)で行った後、インキュベートを検出バッファーに0.175 mg/ml NBTおよび0.35 mg/ml BCIPで行った。反応は、TE, pH 8で終了した。組織切片の対比染色に1 μg/ml DAPIを用い、VECTASHIELD (Vector Laboratory, Burlingame CA) にマウントした。

10

#### 【0190】

##### 膜系ハイブリダイゼーション

1% Sarkosylと1×高リン酸バッファー (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 °Cで2時間、膜をプレ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55 °Cで16時間ハイブリダイズする。ハイブリダイゼーション後、1mM Tris (pH 8.0), 1% Sarkosylで15分間25 °Cで膜を洗浄し、さらに1mM Tris (pH 8.0)で、各回15分間、25 °Cで4回洗浄する。ハイブリダイゼーション複合体を検出するために、XOMAT-ARフィルム (Eastman Kodak, Rochester NY) を膜に一晩70 °Cで暴露し、現像し、視覚で検討する。

20

#### 【0191】

##### ポリマー被覆スライドベースのハイブリダイゼーション

次の方法を、表3に示すマイクロアレイ分析で用いた。プローブを5分間65 °Cに加熱してから、5415C微量遠心機 (Eppendorf Scientific, Westbury NY) で5分間、9400 rpmで遠心分離する。そして18 μlをアレイ表面に等分し、カバーガラスで覆う。アレイは、顕微鏡スライドより僅かに大きい空洞を有する防水チャンバーに移す。チャンバー内を湿度100%に保つため、140 μlの5×SSCをチャンバーの1コーナーに加える。このアレイを含むチャンバーを、60 °Cで約6.5時間インキュベートする。アレイを1×SSC, 0.1% SDS中45 °Cで10分間洗浄し、次に0.1×SSC中45 °Cで10分間の洗浄を3回繰り返した後に、乾燥させる。

30

#### 【0192】

ハイブリダイゼーション反応を絶対的または差次的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に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。差次的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差次的発現を分析する。2つのサンプルのプローブを調製して、異なる標識成分で標識化する。2つの標識されたプローブの混合物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同数のプローブにハイブリダイズされるアレイ上エレメントは、特有の複合蛍光を出す (Shalon W095/35505)。

40

#### 【0193】

50

ハイブリダイゼーション複合体は、Cy3の励起のためには488 nm、Cy5の励起のためには632 nmでスペクトル線を発生し得るInnova 70混合ガス10 Wレーザー (Coherent, Santa Clara CA) を備えた顕微鏡で検出する。励起レーザー光の焦点をアレイ上に置くため、20×顕微鏡対物レンズ (Nikon, Melville NY) を用いる。アレイを含むスライドを顕微鏡のコンピュータ制御のX-Yステージに置き、20 μmの解像度の対物レンズを通してラスタースキャンする。差次的ハイブリダイゼーションフォーマットでは、レーザーにより2つのフルオロフォアを順に励起する。発光された光は波長に基づき分割され、2つのフルオロフォアに対応する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の間に位置するフィルタ群を用いて、シグナルを分離する。用いるフルオロフォアの最大発光の波長は、Cy3では565 nm、Cy5では650 nmである。プローブ混合液に加えた酵母対照mRNAによって生み出されるシグナル強度を用いてスキャンの感度を較正する。アレイ上の或る特定の位置には或る相補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種の重量比1:100,000と相関するようにする。

10

20

30

40

50

#### 【0194】

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した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。デジタル化したデータは、青色(低シグナル)から赤色(高シグナル)までの擬似カラースケールへのリニア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つの異なるフルオロフォアを同時に励起及び測定する場合には、各フルオロフォアの発光スペクトルを用いて、データは先ずフルオロフォア間の光学的クロストーク(発光スペクトルの重なり起因する)について補正する。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が次に統合され、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。

#### 【0195】

##### QPCR分析

QPCRでは、cDNAの合成は、1 μgの総RNAを含む25 μl反応液から行い、これには100単位のM<sup>-</sup>MLV逆転写酵素 (Ambion, Austin TX)、0.5 mM dNTPs (Epicentre, Madison WI)、および40 ng/mlのランダムヘキサマー (Fisher Scientific, Chicago IL) を有した。反応液は25 °Cで10分間、42 °Cで50分間、次に70 °Cで15分間インキュベートし、500 μlに希釈し、30 °Cで保管した。または、正常な組織をClontech (Palo Alto CA) およびClinomicsから購入した。PCRプライマーおよびプローブ(5' 6-FAM標識、3' TAMRA) の設計にPRIME R EXPRESS 1.5ソフトウェア (ABI) を用い、Biosearch Technologies (Novato CA) またはABIで合成した。

#### 【0196】

QPCR反応にはPRISM 7700検出システム (ABI) を用い、これは25 μlの総量に5 μl cDNA鋳型、1x TAQMAN UNIVERSAL PCRマスターミックス (ABI)、100 nMの各PCRプライマー、200 nMのプローブ、また1x VIC標識 -2ミクログロブリン内在性コントロール (ABI) を有した。反応のインキュベートは50 °Cで2分、95 °Cで10分の後、40サイクルのインキュベートを95 °Cで15秒、また60 °Cで1分、行った。発光の測定をサイクルごとに行い、結果の分析にはSEQUENCE DETECTOR 1.7ソフトウェア (ABI) を用い、標準と比較したmRNAの倍単位の差、すなわち相対濃度の計算に、比較C<sub>T</sub>法 (ABI User Bulletin #2) を用いた。QPCRを用いて、図3、4、および5のデータを得た。

#### 【0197】

##### 8 相補的分子

cDNAに相補的なアンチセンス分子であって約5 bpから約5000 bpの長さの分子を用いて、遺伝子発現を検出あるいは阻害する。検出については、実施例7に記載されている。プロ

モータ結合を妨げることにより転写を阻害するためには、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群を含むよう設計する。相補的分子としてはゲノム配列（エンハンサーまたはイントロン等）を含み、これは、ポリメラーゼ、転写因子または調節分子の結合のために充分に開く二重らせんの能力に影響を及ぼす三重らせん塩基対形成に用いられる。翻訳を阻害するためには、相補的分子の設計は、リボソームが該タンパク質をコードするmRNAに結合するのを防ぐようにする。

#### 【0198】

相補的分子を発現ベクターに置いて、効力を試験するため、一過性或いは短期治療向けに器官、腫瘍、滑膜腔また脈管系に、もしくは長期或いは安定した遺伝子治療向けに幹細胞、接合子または他の生殖系統に、細胞株を形質転換するために用いる。非複製ベクターで一過性発現は1ヶ月以上続く、また形質転換/発現系でベクター複製を誘導するエレメントが用いられる場合は3ヶ月以上続く。

相補的分子をコードするベクターを有する分裂する細胞の安定的な形質転換によって、遺伝子導入細胞株、組織または生命体を産生する（米国特許第4,736,866号）。安定した組込みを可能にする充分な量のベクターを同化および複製する細胞も、タンパク質をコードするcDNAの活性を損なうまたは完全に除去するための充分な相補的分子を生成する。

#### 【0199】

##### 9 特異抗体の産生

TMDCのアミノ酸配列の分析にLASERGENEソフトウェア（DNASTAR）を用い、免疫原性の高い領域を決定する。適切なオリゴペプチドを合成し、KLH（Sigma-Aldrich）に抱合する。

#### 【0200】

完全フロイントアジュバント中でオリゴペプチド-KLH複合体を用いてウサギを免疫化し、結果の抗血清の抗ペプチド活性試験を標準ELISA法で行う。抗血清はまた、TMDCの特異的認識について試験する。TMDCで陽性に反応した抗血清の親和性精製をカラムで行い、このカラムはビーズ化アガロース樹脂を含有し、その樹脂に該合成オリゴペプチドを抱合しておく。カラムを平衡化するには12 mLのIMMUNOPURE Gentle Bindingバッファ（Pierce Chemical, Rockford IL）を用いる。3 mLのウサギ抗血清を1 mLの結合バッファと混合し、カラムの上層に加える。カラムの上下にキャップし、抗血清が結合するよう穏やかな振盪を室温で30分行う。カラムを30分間、沈殿させ、自然流下で排液し、16 mL結合バッファ（バッファを4 x 4 mL添加）で洗浄する。抗体の溶出は1 mL分画でIMMUNOPURE Gentle Elutionバッファ（Pierce）を用い、吸光度を280 nmで判定する。ピーク分画をプールし、50 mM Tris, pH 7.4, 100 mM NaCl, および10%グリセロールで透析する。透析後、精製された抗体の濃度判定にBCAアッセイ（Pierce）を用い、アリコット化し、凍結する。

#### 【0201】

##### 10 抗体を用いた免疫精製

天然あるいは組換えタンパク質を精製するには、タンパク質に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で行う。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。該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、該タンパク質が優先的に吸光できるように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、pH2~3のバッファ、或いは高濃度の尿素またはチオシアン酸イオンを用いてカラムから溶出させ、精製されたタンパク質を収集する。

#### 【0202】

##### 11 ウェスタン分析

##### 電気泳動およびプロットティング

タンパク質を有するサンプルを2 xローディングバッファに混合し、95 °Cまで3~5分間、加熱し、4~12% NUPAGE Bis-Trisプレキャストゲル（Invitrogen）にロードする。注記がない限り、等量の総タンパク質を各ウェルにロードする。ゲルの電気泳動は1 x MESまた

10

20

30

40

50

はMOPS泳動バッファ (Invitrogen) 中、200 Vで約45分、Xcell II泳動槽 (Invitrogen) で、RAINBOWマーカー (APB) が分離し色素先端がゲルの底面に近づくまで行う。ゲルと支持体を泳動槽から取って1x転写バッファ (Invitrogen) (10%メタノール含む) 中に数分間、浸漬し、PVDF膜を100%メタノールに数秒間、浸漬してこれを活性化する。膜、ゲル、および支持体をTRANSBLOT SD転写装置 (Biorad, Hercules CA) に置き、定電流350 mAを90分、印加する。

#### 【0203】

##### 抗体とのコンジュゲーションおよび可視化

タンパク質を膜に転写後、膜のブロックは、5% (w/v) 脱脂粉乳を含む1xリン酸緩衝食塩水 (PBS) に0.1% Tween 20界面活性剤 (ブロッキングバッファ) 中、回転振盪機で、少なくとも1時間室温、または4で一晩行う。ブロック後、バッファを取り出し、ブロッキングバッファ内の一次抗体10 mlを加え、インキュベートを回転振盪機で1時間室温、または一晩4で行う。膜の洗浄を3回、各10分PBS-Tween (PBST) で行い、二次抗体を西洋わさびペルオキシダーゼにコンジュゲートしたものを、10 mlブロッキングバッファ中1:3000に希釈して加える。膜と溶液の振盪を30分、室温で行い、次に洗浄を3回、各10分PBSTで行う。

10

#### 【0204】

洗浄溶液を慎重に除去し、膜をECL+化学発光検出システム (APB) で加湿して、約5分インキュベートする。膜を、タンパク質側を下に、BIOMAX Mフィルム (Eastman Kodak) に置き、約30秒、現像する。

20

#### 【0205】

##### 1.2 抗体アレイ

##### タンパク質間相互作用

タンパク質の酵母2ハイブリッド系分析に替えて、抗体アレイを用いてタンパク質間の相互作用とリン酸化を試験できる。種々のタンパク質リガンドを、当該分野で公知の方法で膜に固定する。このアレイのインキュベートを細胞溶解物の存在下で、タンパク質：抗体複合体が形成されるまで行う。当該のタンパク質を同定するため、膜を、当該のタンパク質に特異的な抗体に曝す。または、当該のタンパク質をジゴキシゲニン (DIG) で標識し、膜に曝した後、膜を抗DIG抗体に曝すと、当該のタンパク質が複合体を形成する箇所が表示される。当該のタンパク質と相互作用するタンパク質の同一性は、膜の上の目的タンパク質の位置で判定する。

30

#### 【0206】

##### プロテオームプロファイル

抗体アレイはまた、組換え抗体の高処理スクリーニングに用いる。抗体遺伝子群を有する細菌をロボットで採取し、格子状、高密度 (最大18342の別々の二重スポットしたクローン群) でフィルタに置く。最大15の抗原を一度に用いてクローンをスクリーニングし、結合抗体断片を発現するクローンを同定する。これらの抗体アレイはまた、サンプル中で差次的発現するタンパク質の同定に用い得る (de Wildt, 前出)。

#### 【0207】

##### 1.3 cDNAまたはタンパク質と特異結合する分子のスクリーニング

cDNAまたは其の断片、或いはタンパク質または其の各部分の標識化を、各々<sup>32</sup>P-dCTP, Cy3-dCTPまたはCy5-dCTP (APB) で、或いはBIODIPYまたはFITC (Molecular Probes) で行う。前もって或る基板上に配置した候補分子または化合物のライブラリ群を、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、基板を洗浄し、標識を保持する基板上の、特異結合か複合体形成を示す任意の位置をアッセイし、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、結合した分子との親和性を計算する。

40

#### 【0208】

##### 1.4 2ハイブリッドスクリーン

50

酵母2ハイブリッドシステムであるMATCHMAKER LexAツウハイブリッドシステム(Clontech Laboratories)を用いて、本発明のタンパク質に結合するペプチドをスクリーニングする。該タンパク質をコードする或るcDNAを、pLexAベクターのマルチクローニング部位に挿入し、ライゲーションして、大腸菌に形質転換する。mRNAから調製したcDNAをpB42ADベクターのマルチクローニング部位に挿入し、ライゲーションし、大腸菌に形質転換してcDNAライブラリを作製する。pLexAプラスミドとpB42AD-cDNAライブラリ作製物を大腸菌から単離し、比率2:1で、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。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)培地で平板培養し、30℃でインキュベートしてコロニーを成長させカウントする。コロニーを最小容積1x TE (pH 7.5)でプールし、2%ガラクトース(Gal), 1%ラフィノース(Raf)および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)を補ったSD/-His/-Leu/-Trp/-Ura培地で再プレating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融合タンパク質との相互作用はEGY48内のLEU2レポーター遺伝子の発現を活性化し、ロイシン(-Leu)を含まない培地でコロニー成長を起こす。相互作用は又、X-Galで成長したコロニーに青色を産生する、p8op-lacZレポーター構築物からのガラクトシダーゼの発現を活性化する。

10

#### 【0209】

発現したタンパク質とcDNA融合タンパク質との陽性の相互作用は、個々の陽性コロニーを単離し30℃で1から2日間SD/-Trp/-Ura液体培地で育成することにより検証する。培地のサンプルをSD/-Trp/-Ura培地に蒔き、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のプレートにレプリカプレatingする。ヒスチジンを含まない培地では育成せずヒスチジンを含むSD培地で育成するコロニーは、pLexAプラスミドを失っている。ヒスチジン要求コロニーをSD/Gal/Raf/X-Gal/-Trp/-Uraで育成する。そして白色コロニーを単離し増殖する。該タンパク質と物理的に相互作用する或るタンパク質をコードする或るcDNAを有するpB42AD-cDNAプラスミドを、酵母細胞から単離し特徴付ける。

20

#### 【0210】

##### 1.5 細胞の形質転換アッセイ

##### 軟寒天でのコロニー形成アッセイ

形質転換した細胞の足場非依存性増殖能の測定を、細胞が軟寒天(0.35%)中にコロニーを形成する能力で行う。アッセイは12穴の培養プレートで行い、各ウェルを、固形0.7% Noble寒天(Fisher Scientific, Atlanta GA)を用い細胞増殖培地でコートする。3.5%寒天溶液含有PBSを調製し、高圧滅菌し、マイクロ波をかけ、液状を保つため55℃のウォーターバスで振盪する。寒天の希釈を1:5で0.7%まで、適切な細胞増殖培地で行い、0.5 mlの希釈寒天をプレートの各ウェルに加える。培養プレートを室温に約15分、すなわち寒天が凝固するまで保つ。

30

#### 【0211】

トリプシン処理した細胞の希釈を200~4000細胞/mlまで増殖培地で行い、0.25 mlの希釈細胞を2 mlの温かな0.35%寒天と混ぜる。希釈した細胞を培養プレートのウェルに加え、二つ組のウェルを各細胞濃度について準備する。プレートを冷ますため約30分、室温に置いた後、37℃のインキュベータに移す。1~2週間のインキュベート後、コロニーのカウントを倒立位相差顕微鏡で行う。コロニー形成効率の判定を、百分率の形成コロニー/総培養細胞数として行う。

40

#### 【0212】

##### アポトーシス/生存アッセイ

形質転換した細胞がアポトーシス(プログラム細胞死)を回避し延命する能力の測定は、或るアッセイで、培養細胞のアポトーシスまたは延命をFACS分析で二重染色法にAnnexin Vとヨウ化プロピジウム(PI)を用いて判定して行える。Annexin Vはアポトーシス細胞用マーカーとして働き、これはアポトーシスの細胞表面マーカーであるホスファチジルセリ

50

ンへの結合による。PIとの対比染色により、Annexin V陽性でPI陰性のアポトーシス細胞と、Annexin VおよびPIに陽性の壊死細胞を区別できる。アポトーシスの測定は培養0～24時間で行い、細胞生存の測定は培養24～96時間で行う。

【0213】

或いは、分泌タンパク質たとえばHUPAPによるアポトーシス/細胞生存への直接効果の測定を、培養したヒト血管内皮細胞(HMVEC)で、HMVEC細胞をHUPAPで処理して、または、細胞を、HUPAPをコードするcDNAを有する組換えアデノウイルスに感染させて行える。HMVEC細胞のアポトーシス/生存は、上記のように測定する。

【0214】

組織浸潤および転移のアッセイ

癌化した腫瘍細胞による細胞遊走と組織浸潤の判定には、BICOAT Angiogenesisシステム(BD Biosciences, Franklin Lakes NJ)を、製造者の記載のように用いる。アッセイは、BD FALCONマルチウェル インサートプレートに、再構成BD MATRIGEL基底膜マトリクスで均一に被覆した8 μm孔径のBD FLUOROBLOKポリエチレンテレフタレート膜を入れ、無処理マルチウェル レシーバープレートに挿入して行う。本システムは細胞の、膜を通じた受動拡散への障壁をもたらすが、浸潤癌細胞による能動遊走を許容する。細胞を適切な培地に入れ、インキュベートをチャンバーの上部で好適な時間おこなった後、膜の下側に現れる細胞を定量する。膜は490～700nmの光の透過を防ぐので、膜を横断する細胞の検出を、細胞数に比例する蛍光で行う。

【0215】

本明細書で言及した全ての特許及び刊行物は、参照により開示に含まれる。本発明の範囲および趣旨から逸脱しない限りでの、記載した本発明の方法およびシステムの、種々の修正および変更は、当業者には自明であろう。本発明について記載するにあたり特定の好適実施態様に関連して説明を行ったが、請求される本発明が、そのような特定の実施態様に不当に限定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、本明細書に記載した本発明の実施方法の様々な変更は、特許請求の範囲内にあることを意図するものである。

【0216】

(表の簡単な説明)

表(Table)1は、LIFESEQ Goldデータベース(Incyte Genomics, Palo Alto CA)を使用して作製したTMDCのノーザン分析を示す。列1は組織カテゴリ、列2は組織カテゴリのクローンの数、列3は少なくとも1つの転写物が見られたライブラリの数と該カテゴリのライブラリの総数、列4は転写物の絶対存在量(転写物の数)、列5は転写物の存在%を示す。

【0217】

表2はTMDCのノーザン分析を示すが、これは消化系組織であって転写物が過剰発現する、すなわち存在量(Abundance)が1より大きい転写物が何れか1つのcDNAライブラリ内で見られる組織での分析である。列1はライブラリID(訳注:次列はcDNA数)、列2はライブラリの説明、列3は絶対存在量(転写物/ライブラリの数)、列4は転写物の存在%を示す。

【0218】

表3は正常な大腸組織と比較した、大腸癌を有する患者の組織でのTMDCの差次的発現を示し、これはマイクロアレイ分析で判定した。列1は腫瘍サンプルと正常な組織の間の差次的発現(DE)を示す。結果は腫瘍/正常発現の比という形で表現する。列2(P1説明)は、緑色蛍光色素Cy3で標識した微視的正常サンプルの、組織と患者供与者(Dn)を示す。列3(P2説明)は、赤色蛍光色素Cy5で標識した病変サンプル(大腸腫瘍または大腸ポリープ)の組織と患者供与者(Dn)を示す。

【図面の簡単な説明】

【0219】

【図1-A】図は膜貫通タンパク質腫瘍抗原(TMDC、SEQ ID NO:1)を示し、これはcDNA

10

20

30

40

50

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

【図1 - B】図は膜貫通タンパク質腫瘍抗原 (TMDC、SEQ ID NO:1) を示し、これはcDNA (SEQ ID NO:2) によってコードされる。アラインメントはMACDNASIS PROソフトウェア (日立ソフトウェアエンジニアリング, South San Francisco CA) を用いて作成した。

【図1 - C】図は膜貫通タンパク質腫瘍抗原 (TMDC、SEQ ID NO:1) を示し、これはcDNA (SEQ ID NO:2) によってコードされる。アラインメントはMACDNASIS PROソフトウェア (日立ソフトウェアエンジニアリング, South San Francisco CA) を用いて作成した。

【図1 - D】図は膜貫通タンパク質腫瘍抗原 (TMDC、SEQ ID NO:1) を示し、これはcDNA (SEQ ID NO:2) によってコードされる。アラインメントはMACDNASIS PROソフトウェア (日立ソフトウェアエンジニアリング, South San Francisco CA) を用いて作成した。

【図1 - E】図は膜貫通タンパク質腫瘍抗原 (TMDC、SEQ ID NO:1) を示し、これはcDNA (SEQ ID NO:2) によってコードされる。アラインメントはMACDNASIS PROソフトウェア (日立ソフトウェアエンジニアリング, South San Francisco CA) を用いて作成した。

【図1 - F】図は膜貫通タンパク質腫瘍抗原 (TMDC、SEQ ID NO:1) を示し、これはcDNA (SEQ ID NO:2) によってコードされる。アラインメントはMACDNASIS PROソフトウェア (日立ソフトウェアエンジニアリング, South San Francisco CA) を用いて作成した。

【図1 - G】図は膜貫通タンパク質腫瘍抗原 (TMDC、SEQ ID NO:1) を示し、これはcDNA (SEQ ID NO:2) によってコードされる。アラインメントはMACDNASIS PROソフトウェア (日立ソフトウェアエンジニアリング, South San Francisco CA) を用いて作成した。

【図1 - H】図は膜貫通タンパク質腫瘍抗原 (TMDC、SEQ ID NO:1) を示し、これはcDNA (SEQ ID NO:2) によってコードされる。アラインメントはMACDNASIS PROソフトウェア (日立ソフトウェアエンジニアリング, South San Francisco CA) を用いて作成した。

【図2】図2は、TMDCの疎水性プロットを示す。負のY軸が疎水性を示し、X軸はアミノ酸残基番号の位置/数である。プロット作成にはMACDNASIS PROソフトウェアを用いた。

【図3】図3は多様な正常な成体組織でのTMDCの発現を示す。X軸は組織のタイプ、Y軸は正常な大腸組織に見られる発現 (100%に設定) に比較したTMDCの発現を示す。QPCR分析はTAQMANプロトコル (Applied Biosystems (ABI), Foster City CA) によって実施した。組織はClinomics (Pittsfield MA) およびClontech (Palo Alto CA) から得た。分析に用いたオリゴヌクレオチドプローブは、SEQ ID NO:2のほぼヌクレオチド1899からほぼヌクレオチド1966に及ぶ。

【図4】図4は、QPCR (ABI) でドナーをマッチさせた正常な大腸組織と比較した、大腸癌患者の組織内のTMDCの差次的発現を示す。X軸は患者ID (ドナーID)、Y軸は正常な大腸組織に見られる発現 (100%に設定) に比較したTMDCの発現を示す。腫瘍サンプルは黒で表示され、正常な組織は白で表示される。分析に用いたオリゴヌクレオチドプローブは、SEQ ID NO:2のほぼヌクレオチド1899からほぼヌクレオチド1966に及ぶ。

【図5】図5は、QPCR (ABI) を用いて或る非腫瘍化大腸細胞株 (LS123) 及び正常な大腸組織に見られる発現 (100%に設定) と比較した、種々の大腸腫瘍細胞株内のTMDCの差次的発現を示す。細胞株はATCC (Manassas VA) から得た。分析に用いたオリゴヌクレオチドプローブは、SEQ ID NO:2のほぼヌクレオチド1899からほぼヌクレオチド1966に及ぶ。

【図6】図6は、正常な大腸組織でのTMDCをコードする転写物の発現を示す。薄切片はDAP I染色し、その*in situ*ハイブリダイズに用いたセンスまたはアンチセンスRNAプローブはSEQ ID NO:2の或る断片から作製し、断片はSEQ ID NO:2のほぼヌクレオチド1068からほぼヌクレオチド2324に及ぶ。

【図7】図7は、或る大腸絨毛腺癌での、TMDCをコードする転写物の発現を示す。薄切片はDAPI染色し、その*in situ*ハイブリダイズに用いたアンチセンスRNAプローブはSEQ ID NO:2の或る断片から作製し、断片はSEQ ID NO:2のほぼヌクレオチド1068からほぼヌクレオチド2324に及ぶ。

【0220】

【表1】

10

20

30

40

50

表 1

組織カテゴリー	cDNA数	ライブラリ数	存在量	%存在比
心血管系	272986	1/74	1	0.0004
結合織	151678	0/54	0	0.0000
消化器系	521762	19/155	40	0.0077
胎児構造	108468	0/24	0	0.0000
内分泌系	233683	0/63	0	0.0000
外分泌腺	258383	5/67	7	0.0027
生殖器、女性	456353	5/117	7	0.0015
生殖器、男性	463016	12/120	13	0.0028
性細胞	48181	0/5	0	0.0000
血液および免疫系	725942	1/179	1	0.0001
肝臓	115620	1/37	1	0.0009
筋骨格系	162801	0/50	0	0.0000
神経系	995533	0/231	0	0.0000
脾臓	111771	2/25	2	0.0018
呼吸器系	412898	7/96	9	0.0022
感覚器	25345	0/10	0	0.0000
皮膚	72732	0/18	0	0.0000
顎口腔系	14712	0/17	0	0.0000
未分類/混合性	159180	4/22	4	0.0025
尿路	295517	2/68	2	0.0007
計	5606561	59/1432	87	0.0000

10

20

30

【 0 2 2 1 】

【 表 2 】

表 2

ライブラリID*	cDNA数	ライブラリの説明	存在量	%存在比
COLCTUT03	3344	大腸腫瘍、盲腸、腺癌、70女	3	0.0897
COLNTUT03	5063	結腸腫瘍、S状結腸、腺癌、62男、m/COLNNOT16	3	0.0593
COLCDIT01	3508	大腸、盲腸、良性家族性ポリポーシス、16男	2	0.0570
COLRTUE01	3657	大腸腫瘍、直腸、腺癌、管状腺腫とマッシュ、50男、SRP	2	0.0547
SINIDME01	4188	小腸、回腸、慢性炎症、29女、SRP、EF	2	0.0478
STOMTUP02	18951	胃腫瘍、腺癌、低分化、3' CGAP	8	0.0422
COLNTUP17	7421	大腸腫瘍、腺癌、3' CGAP	2	0.0270

\* 全てのノーマライズまたはプールのライブラリ、混合組織ライブラリ、および細胞株は、この分析から除外した。

【 0 2 2 2 】  
【 表 3 】

10

20

30

40

表 3

DE	P1の説明	P2の説明	
1.5	ヒト、大腸、Nrml, 腺癌とマッチ, Dn3579	ヒト、大腸腫瘍、腺癌, Dn3579	
1.9	ヒト、大腸、Nrml, 腺癌とマッチ, Dn3839	ヒト、大腸腫瘍、腺癌, Dn3839	
2.2	ヒト、大腸、Nrml, 腺癌とマッチ, Dn3839	ヒト、大腸腫瘍、腺癌, Dn3839	
			10
			20
			30
			40

【 1 - A 】

5' ATG CTG TCC GAT GAC CAC GTG AAT GAA ATC ATC ATA CAG GTT GAG AAT GTT TCC 54  
 M L S D D H V N E I I I Q V E N V S  
 TCT GGG GTC CAA AGC CAC CCA TCC TCA AAT CAG AAT TTT CAA GAA AAG GTG CTG 108  
 S G V Q S H P S S N Q I F Q E K V L  
 CTA GAC TCA AGC ATC AAC ATG GTT TTG TCA ATA TCT GAC AAT GAT GTG ATA GAC 162  
 L D S S I N M V L S I S D I D V I D  
 TCT CAG ACA GTC AGC AAA AGG AAT GAC CAA AAG GGT AAC CAG GTG CTG CGG TTT 216  
 S Q T V S K R N D Q K G N Q V L R F  
 TCA ACA TCT TTG AAT GAG TCG ATG TCT CAG ACC CTT CNT ACC CTA GAA TGC ATG 270  
 S T S L N E S M S Q T L H S L E C M  
 GGC ATA GAC ACT CCT GGT TCT TCA CAT GAA ACT GTT CAA GGA CAG AAG TTA ATC 324  
 G I D T P G S S H E T V Q G Q K L I  
 GCA TCC CTT ATA CCC ATG ACA TCC AGA GAC AGA AAT AAA GCC ATC AGG AAC CAG 378  
 A S L I P M T S R D R I K A I R N Q

【 1 - B 】

CCA AGG ACC ATG GAA GAG AAA AGG AAC CTT AGG AAA ATA GTT GAC AAA GAA AAA 432  
 P R T M E E K R N L R K I V D K E K  
 AGC AAA CAG ACC CAT CTT ATC CTT CAG CTC AAT TGC TGT AAT CAG TGT CTG AAC 486  
 S K Q T H R I L Q L N C C I Q C L N  
 TCC AAT TCC GGG GCT TAT CCG AGA TCC AGG AAC AGC CTG TCG GAA AAT CTG AAT 540  
 S I S R A Y R R S K N S L S E I L N  
 TCC ATC ACC CTG TCG CAG AAG ACG CTG AAG ATC ATT GGA GCC AAG TTT GGA ACC 594  
 S I S L W Q K T L K I I G K F G I  
 AGC GTC CTC TCC TAT TTC AAC TTT CTG AGA TGG CTT TTG AAG TTC AAC AAT TTC 648  
 S V L S Y F N F L R W L L K F N I F  
 TCA TTC ATC CTG AAC TTC AGC TTC ATC ATA ATC CCT CAG TTT ACC GTG GCC AAA 702  
 S F I L N F S F I I I P Q F T V A K  
 AAG AAC ACC CTC CAG TTC ACT GGG CTG GAG TTT TTC ACT GGG GGT TAT TTT 756  
 K N T L Q F T G L E F F T G V Y F

【 1 - C 】

AGG GAC ACA GTG ATG TAC TAT GGC TTT TAC ACC AAT TCC ACC ATC CAG CAC GGC 810  
 R D T V M Y Y G F Y T N S T I Q H G  
 AAC AGC GGC GCA TCC TAC AAC ATG CAG CTG GCC TAC ATC TTC ACA ATC GGA GCA 864  
 N S G A S Y N M Q L A Y I F T I G A  
 TGC TTG ACC ACC TGC TTC AGT TTG CTG TTC GGC AAG TAT TTC CGS 918  
 C L T T C F F S L L F S M A K Y F R  
 AAC AAC TTC AAT AAT CCC CAC AAT TAC TCC GGA GGG ATC ACC AAG CTG ATC TTT 972  
 N N F I N P H I Y S G I T K L I F  
 TGC TGG GAC TTC ACT GTC ACT CAT GAA AAA GCT GTG AAG CTA AAA CAG AAG AAT 1026  
 C W D F T V T H E K A V K L K Q K N  
 CTT AGC ACT GAG ATA AGG GAG AAC CTG TCA GAG CTC CGT CAG GAG AAT TCC AAG 1080  
 L S T E I R E N L S E L R Q E N S K  
 TTG ACG TTC AAT CAG CTG ACC CCG TTC TCT GCC TAC ATG GTA GCC TGG GTT 1134  
 L T F N Q L L T R F S A Y M V A W V

【 1 - D 】

GTC TCT ACA GGA GTG GCC ATA GCC TGC TGT GCA GCC GTT TAT TAC CTG GCT GAG 1188  
 V S T G V A I A C C A A V Y Y L A E  
 TAC AAC TTA GAG TTC CAG ACA AGT AAC CCT GGG GCG GTG CTG TTA CTG 1242  
 Y N L E F L K T H S N P G A V L L L  
 CCT TTC GTT GTG TCC TGC AAT AAT CTG GCC GTG CCA TGC ATC TAC TCC ATG TTC 1296  
 F F V V S C I N L A V P C I Y S M F  
 AGG CTT GTG GAG AGG TAC GAG ATG CCA CCG CAC GMA GTC TAC GTT CTC CTG ATC 1350  
 R L V E R Y E M P R H E V Y V L L I  
 CGA AAC ATC TTT TTG AAA ATA TCA ATC AAT CTT TGT TAC TAT TGG CTC 1404  
 R N I F L K I S I I G I L C Y Y W L  
 AAC ACC GTG GCC CTG TCT GGT GAA GAG TGT TGG GAA ACC CTC AAT GGC CAG GAC 1458  
 N T V A L S G E E C W E T L I G Q D  
 ATC TAC CCG CTC CTT CTG ATG GAT TTT GTG TTC TCT TTA GTC AAT TCC TTC CTG 1512  
 I Y R L L L M D F V F S L V N S F L

【 ☒ 1 - F 】

1521 1530 1539 1548 1557 1566  
GGG GAG TTT CTG AGG AGA ATC AAT GGG ATG CAA CTG ATC ACA AGT CTT GCC CTT  
G E F L R R I I G M Q L I T S L G L

1575 1584 1593 1602 1611 1620  
CAG GAG TTT GAC AAT GCC AGG AAC GTT CTA GAA CTG ATC TAT GCA CAA ACT CTG  
Q E F D I A R N V L E L I Y A Q T L

1629 1638 1647 1656 1665 1674  
GTG TGG AAT GGC ATC TTC TGC CCC CTG CTG CCC TTT ATC CAA ATG AAT ATG  
V W I G I F F C P L L P F I Q M I M

1683 1692 1701 1710 1719 1728  
CTT TTC ATC ATG TTC TAC TCC AAA AAT ATC ASC CTG ATG ATG AAT TTC CAG CCT  
L F I M F Y S K N I S L M M N F Q P

1737 1746 1755 1764 1773 1782  
CCG AGC AAA GCC TGG CCG GCC TCA CAG ATG ATG ACT TTC TTC ATC TTC TTG CTC  
P S K A W R A S Q M M T F F I F L L

1791 1800 1809 1818 1827 1836  
TTT TTC CCA TCC TTC ACC GGG GTC TTG TGC ACC CTG GCC ATC ACC ATC TGG AGA  
F F P S F T G V L C T L A I T I W R

1845 1854 1863 1872 1881 1890  
TTG AAG CCT TCA GCT GAC TGT GGC CCT TTT CGA GGT CTG CCT CTC TTC AAT CAC  
L K P S A D C G P F R G L P L F I H

【 ☒ 1 - F 】

1899 1908 1917 1926 1935 1944  
TCC ATC TAC AGC TGG ATC GAC ACC CTA AGT ACA CCG CCT GGC TAC CTG TGG GTT  
S I Y S W I D T L S T R P G Y L W V

1953 1962 1971 1980 1989 1998  
GTT TGG ATC TAT CCG AAC CTC AAT GGA AGT GTG CAC TTC TTT TTC ATC CTC ACC  
V W I Y R N L I G S V H F F I L T

2007 2016 2025 2034 2043 2052  
CTC AAT GTG CTA ATC ATC ACC TAT CTT TAC TGG CAG ATC ACA GAG GGA AGG AAG  
L I V L I I T Y L Y W Q I T E G R K

2061 2070 2079 2088 2097 2106  
AAT ATG ATA AGG CTG CTC CAT GAG CAG ATC AAT AAT GAG GGC AAA GAT AAA ATG  
I M I R L L H E Q I I N E G K D K M

2115 2124 2133 2142 2151 2160  
TTC CTG ATA GAA AAA TTG ATC AAG CTG CAG GAT ATG GAG AAG AAA GCA AAC CCC  
P L I E K L I K L Q D M E K K A N P

2169 2178 2187 2196 2205 2214  
AGC TCA CTT GTT CTG GAA AGG AGA GAG GTG GAG CAA CAA GGC TTT TTG CAT TTG  
S S L V L E R R E V E Q Q G F L H L

2223 2232 2241 2250 2259 2268  
GGG GAA CAT GAT GGC AGT CTT GAC TTG CGA TCT AGA AGA TCA GTT CAA GAA GGT  
G E H D G S L D L R S R R S V Q E G

【 ☒ 1 - G 】

2277 2286 2295 2304 2313 2322  
AAT CCA AGG GCC TGA TGA CTC TTT TGG TAA CCA GAC ACC AAT CAA ATA AGG GGA  
N P R A

2331 2340 2349 2358 2367 2376  
GGA GAT GAA AAT GGA ATG AAT TCT TCC ATG CCA CCT GTG CCT TTA GGA ACT GCC

2385 2394 2403 2412 2421 2430  
CAG AAG AAA ATC CAA GGC TTT AGC CAG GAG CCG AAA CTG ACT ACC ATG TAA TTA

2439 2448 2457 2466 2475 2484  
TCA AAG TAA AAT TGG GCA TTC CAT GCT AAT TTT AAT ACC TGG ATT GCT GAT TTT

2493 2502 2511 2520 2529 2538  
TCA AGA CAA AAT ACT TGG GGT TTT CCA ATA AAG AAT GTT GTA ATA TTG AAA TGA

2547 2556 2565 2574 2583 2592  
GCC TAC AAA AAC CTA GGA AGA GAT AAC TAG GGA ATA ATG TAT AAT ATC TTC AAG

2601 2610 2619 2628 2637 2646  
AAA TGT GTG CAG GRA TGA TTG GTT CTT AGA AAT CTC TCC TGC CAG ACT TCC CAG

2655 2664 2673 2682 2691 2700  
ACC TGG CAA AGG TTT AGA AAC TGT TGC TAA GAA AAG TGG TCC ATC CTG AAT AAA

2709 2718 2727 2736 2745 2754  
CAT GTA ATA CTC CAG CAG GGA TAT GAA GCC TCT GAA TTG TAG AAC CTG CAT TTA

【 ☒ 1 - H 】

2763 2772 2781 2790 2799 2808  
TTT GTG ACT TTG AAC TAA AGA CAT CCC CCA TGT CCC AAA GGT GGA ATA CAA CCA

2817 2826 2835 2844 2853 2862  
GAG GTC TCA TCT CTG AAC TTT CTT GCG TAC TGA TTA CAT GAG TCT TTG GAG TCG

2871 2880 2889 2898 2907 2916  
GGG ATG GAG GAG GTT CTG CCC CTG TGA GGT GTT ATA CAT GAC CAT CAA AGT CCT

2925 2934 2943 2952 2961 2970  
ACG TCA AGC TAG CTT TGC ACA GTG GCA GTA CCG TAG CCA ATG AGA TTT ATC CGA

2979 2988 2997 3006 3015 3024  
GAC GCG AAT AAT GCT AAT TGG AAA TTT TCC CAA TAC CCC ACC GTG ATG ACT TGA

3033 3042 3051 3060 3069 3078  
AAT ATA ATC AGC GCT GGC AAT TTT TGA CAG TCT CTA CCG AAG CTG AAT AAG AAA

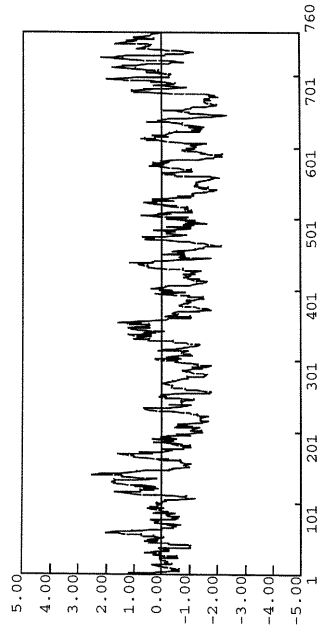
3087 3096 3105 3114 3123 3132  
AAA GAA AAG AAA AGA AAT TGG CTG GGT CCG ATG GCT TAT GCC TGT AAT CCC GGC

3141 3150 3159 3168 3177 3186  
ACT TTG GGA GGC TGA GGC AAG CCG ATC ACT TAA TGT CAG GAG TTC AAG ACC AGC

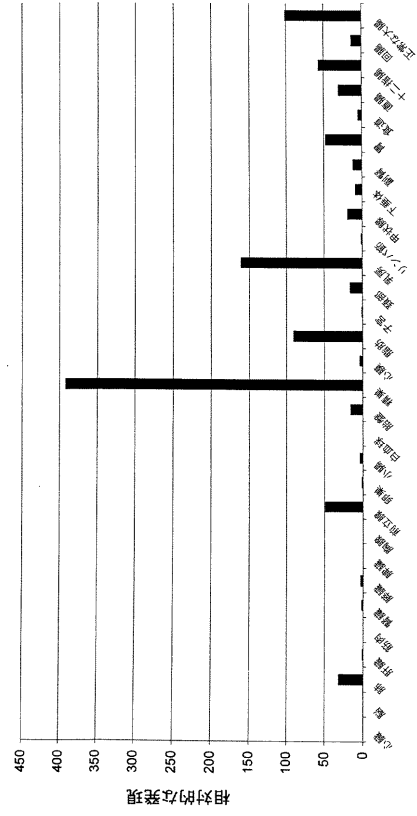
3195 3204 3213 3222 3231 3240  
CTG GCC AAC ATG GTG AAA CCC CGT CTC TAC TAA GGA TAA AAA AAC TGG CTG GGC

3249  
GTG GTG GTA CAT GCC T 3

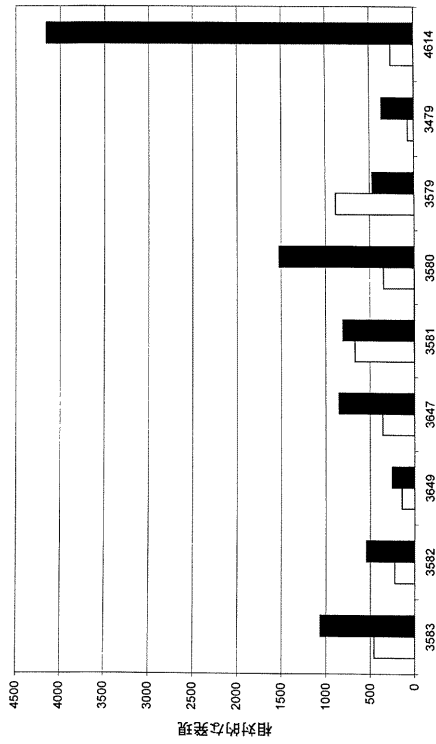
【 图 2 】



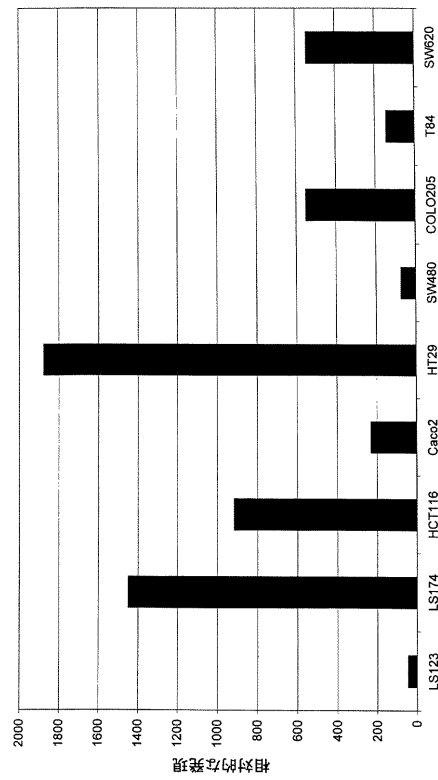
【 图 3 】



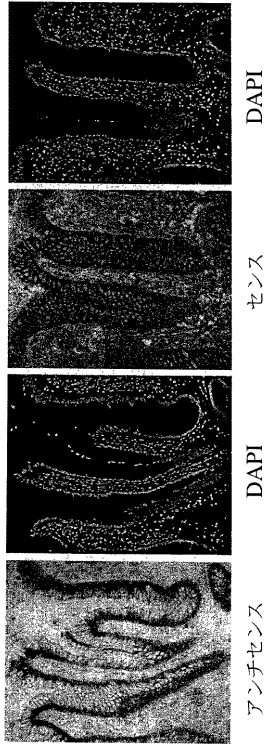
【 图 4 】



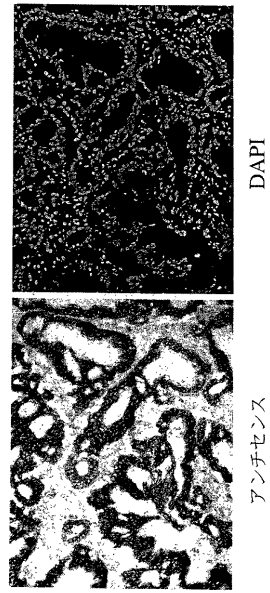
【 图 5 】



【 図 6 】



【 図 7 】



【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  
6 March 2003 (06.03.2003)

PCT

(10) International Publication Number  
WO 03/018768 A2

- (51) International Patent Classification: C12N Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US); BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US).
- (21) International Application Number: PCT/US02/27144
- (22) International Filing Date: 22 August 2002 (22.08.2002) (74) Agents: STREETER, David, G. et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/314,914 24 August 2001 (24.08.2001) US
- (71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (72) Inventors: and (75) Inventors/Applicants (for US only): TRIBOULEY, Catherine, M. [FR/US]; 1121 Tennessee St., # 5, San Francisco, CA 94107 (US). LASEK, Amy, K. W. [US/US]; 297 41st #5, Oakland, CA 94611 (US). YUE, (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).

[Continued on next page]

(54) Title: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER

```

          9           18           27           36           45           54
5' ATG CTG TCC GAT GAC CAC GTG AAT GAA ATC ATC ATA CAG GTT CAG AAT GTT TCC
   M  L  S  D  D  H  V  N  E  I  I  E  Q  V  E  N  V  S

          63           72           81           90           99           108
TCT GAG GTC CAA AGC CAC CCA TCC TCA AAT CAG ATT TTT CAA GAA AAG GTG CTG
   S  G  V  Q  S  H  P  S  S  N  Q  I  F  Q  E  R  V  L

          117          126          135          144          153          162
CTA GAC TCA AGC ATC AAC ATG GTT TTG TCA ATA TCT GAC ATT GAT GTG ATA GAC
   L  D  S  S  I  K  M  V  L  S  I  E  D  I  D  V  I  D

          171          180          189          198          207          216
TCT CAG ACA GAC AGC AAA AGG AAT GAC CAA AAG GGT AAC CAG GTG CTG CGG TTT
   S  Q  T  V  S  K  R  N  D  Q  K  G  N  Q  V  L  R  F

          225          234          243          252          261          270
TCA ACA TCT TTG AAT GAG TCG ATG TCT CAG ACC CTT CAG AGC CTA GAA TCC ATG
   S  T  S  L  N  E  S  M  S  Q  T  L  H  S  L  E  C  M

          279          288          297          306          315          324
GGC ATA GAC ACT CCT GGT TCT TCA CAG GAA ACT GTT CAA GGA CAG AAG TTA ATC
   G  I  D  T  P  G  S  S  H  E  T  V  Q  G  Q  K  L  I

          333          342          351          360          369          378
GCA TCC CTT ATA CCC ATG ACA TCC AGA GAG AGA ATT AAA GCC ATC AGG AAC CAG
   A  S  L  I  P  Y  T  S  R  D  R  I  K  A  I  R  N  Q

```

(57) Abstract: The invention provides a transmembrane protein, TMDC, that is differentially expressed in a colon or stomach cancer. It also provides for the use of the protein, a cDNA encoding the protein, and antibodies that specifically bind the protein in various methods to diagnose, stage, treat, or monitor the progression or treatment of a colon or stomach cancer.

WO 03/018768 A2

**WO 03/018768 A2** 

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

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

WO 03/018768

PCT/US02/27144

**TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER****TECHNICAL FIELD**

5 This invention relates to a transmembrane protein differentially expressed in cancer, its encoding cDNA, and an antibody that specifically binds the protein and to their use to diagnose, to stage, to treat, or to monitor the progression or treatment of colon or stomach cancer.

**BACKGROUND OF THE INVENTION**

10 Array technologies and quantitative PCR provide the means to explore the expression profiles of a large number of related or unrelated genes. When an expression profile is examined, arrays provide a platform for examining which genes are tissue-specific, carrying out housekeeping functions, parts of a signaling cascade, or specifically related to a particular genetic predisposition, condition, disease, or disorder. The application of expression profiling is particularly relevant to improving diagnosis, prognosis, and treatment of the disease. For example, both the sequences and  
15 the amount of expression can be compared between tissues from subjects with different types of cancer.

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

Transmembrane proteins (TM), e.g., proteins which traverse a cell membrane, are both potential markers and therapeutic targets for a disease condition. For example, if associated with a tumor cell, many TM proteins act as cell-surface receptors involved in signal transduction pathways that control growth and differentiation in cells. Thus in a disease state, modulation of TM activity or  
25 function may interfere with the disease process.

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

Several molecular pathways have been linked to the development of colon cancer, and the  
35 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

WO 03/018768

PCT/US02/27144

provide an early indicator of colon cancer or a predisposition for the development of colon cancer. These proteins can also be used as therapeutic targets to identify molecules useful for treatment of cancer.

A number of genes associated with the predisposition, development, and progression of colon cancer have been identified. For example, it is well known that abnormal patterns of DNA methylation occur consistently in human tumors. In colon cancer in particular, it has been found that these changes occur early in tumor progression; for example, in premalignant polyps that precede colon cancer. DNA methyltransferase, the enzyme that performs DNA methylation, is significantly increased in histologically normal mucosa from patients with colon cancer or the benign polyps that precede cancer, and this increase continues during the progression of colonic neoplasms (Wafik *et al.* (1991) *Proc Natl Acad Sci* 88:3470-3474).

Familial Adenomatous Polyposis (FAP) is a rare autosomal dominant syndrome that precedes colon cancer and is caused by an inherited mutation in the adenomatous polyposis coli (APC) gene. The APC gene is a part of the APC- $\beta$ -catenin-Tcf (T-cell factor) pathway. Impairment of this pathway results in the loss of orderly replication, adhesion, and migration of colonic epithelial cells and in the growth of polyps. Hereditary Nonpolyposis Colorectal Cancer (HNPCC) is another inherited autosomal dominant syndrome that is distinguished by the tendency to early onset of colon cancer and the development of other cancers. HNPCC results from the mutation of one or more genes in the DNA mis-match repair (MMR) pathway. Mutations in two human MMR genes, MSH2 and MLH1, are found in a large majority of HNPCC families identified to date. Almost all colon cancers arise from cells in which the estrogen receptor (ER) gene has been silenced. The silencing of ER gene transcription is age related and linked to hypermethylation of the ER gene (Issa *et al.* (1994) *Nature Genet* 7:536-540). Introduction of an exogenous ER gene into cultured colon carcinoma cells results in marked growth suppression.

Clearly there are a number of genetic alterations associated with colon cancer and with the development and progression of the disease that potentially provide early indicators of cancer development. These alterations may be monitored and perhaps corrected therapeutically.

The discovery of a transmembrane protein, its encoding cDNA, and the making of an antibody that specifically binds the protein satisfies a need in the art by providing compositions which are useful to diagnose, to stage, to treat, or to monitor the progression or treatment of a colon or stomach cancer.

#### SUMMARY OF THE INVENTION

The invention is based on the discovery of a transmembrane protein differentially expressed in cancer that has been designated TMDC, its encoding cDNA, and an antibody that specifically binds the protein. These molecules are useful to diagnose, to stage, to treat, or to monitor the

WO 03/018768

PCT/US02/27144

progression or treatment of a colon or stomach cancer.

The invention provides an isolated cDNA comprising a nucleic acid sequence encoding a protein having the amino acid sequence of SEQ ID NO:1. The invention also provides an isolated cDNA or the complement thereof selected from a nucleic acid sequence of SEQ ID NO:2; a fragment of SEQ ID NO:2 selected from SEQ ID NOs:3-10, and a variant of SEQ ID NO:2 selected from SEQ ID NOs:12-16. The invention further provides a probe consisting of the cDNA encoding the transmembrane protein,

A cell transformed with the cDNA encoding the transmembrane protein, a composition comprising the cDNA encoding the transmembrane protein, and a labeling moiety, an array element comprising the cDNA encoding the transmembrane protein, and a substrate upon which the cDNA encoding the transmembrane protein, is immobilized.

The invention provides a vector containing the cDNA encoding TMDC, a host cell containing the vector and a method for using the cDNA to make the protein, the method comprising culturing the host cell containing the vector containing the cDNA encoding the protein under conditions for expression and recovering the protein from the host cell culture. The invention also provides a transgenic cell line or organism comprising the vector containing the cDNA encoding TMDC. The invention further provides a composition, a substrate or a probe comprising the cDNA, a fragment, a variant, or complements thereof, which can be used in methods of detection, screening, and purification. In one aspect, the probe is a single-stranded complementary RNA or DNA molecule.

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

The invention provides a method for using a cDNA 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 to allow specific binding and detecting specific binding to the cDNA, thereby identifying a ligand which specifically binds the cDNA. In one aspect, the molecules or compounds are selected from antisense molecules, artificial chromosome constructions, branched nucleic acids, DNA molecules, enhancers, peptides, peptide nucleic acids, proteins, RNA molecules, repressors, and transcription factors. The invention also provides a method for using a cDNA to purify a ligand which specifically binds the

WO 03/018768

PCT/US02/27144

cDNA, the method comprising attaching the cDNA to a substrate, contacting the cDNA with a sample under conditions to allow specific binding, and dissociating the ligand from the cDNA, thereby obtaining purified ligand. The invention further provides a method for assessing efficacy or toxicity of a molecule or compound comprising treating a sample containing nucleic acids with the molecule or compound; hybridizing the nucleic acids with a cDNA under conditions for hybridization complex formation; determining the amount of complex formation; and comparing the amount of complex formation in the treated sample with the amount of complex formation in an untreated sample, wherein a difference in complex formation indicates the efficacy or toxicity of the molecule or compound.

10 The invention provides a purified protein or a portion thereof selected from the group consisting of an amino acid sequence of SEQ ID NO:1, an antigenic epitope of SEQ ID NO:1, and a variant OF SEQ ID NO:1 having at least 90% amino acid sequence identity to the amino acid sequence of SEQ ID NO:1. The invention also provides a composition comprising the purified protein and a pharmaceutical carrier, a composition comprising the protein and a labeling moiety, a substrate upon which the protein is immobilized, and an array element comprising the protein. The invention further provides a method for detecting expression of a protein having the amino acid sequence of SEQ ID NO:1 in a sample, the method comprising performing an assay to determine the amount of the protein in a sample; and comparing the amount of protein to standards, thereby detecting expression of the protein in the sample. The invention still further provides a method for 20 diagnosing cancer comprising performing an assay to quantify the amount of the protein expressed in a sample and comparing the amount of protein expressed to standards, thereby diagnosing a neoplastic disorder. In a one aspect, the assay is selected from antibody arrays, enzyme-linked immunosorbent assays, fluorescence-activated cell sorting, 2D-PAGE and scintillation counting, protein arrays, radioimmunoassays, and western analysis. In a second aspect, the sample is selected 25 from colon or stomach tissue. In a third aspect, the cancer is a colon or stomach cancer.

The invention provides a method for using a protein to screen a library or a plurality of molecules or compounds to identify at least one ligand, the method comprising combining the protein with the molecules or compounds under conditions to allow specific binding and detecting specific binding, thereby identifying a ligand which specifically binds the protein. In one aspect, the 30 molecules or compounds are selected from agonists, antagonists, bispecific molecules, DNA molecules, small drug molecules, immunoglobulins, inhibitors, mimetics, multispecific molecules, peptides, peptide nucleic acids, pharmaceutical agent, proteins, and RNA molecules. In another aspect, the ligand is used to treat a subject with a neoplastic disorder. The invention also provides an therapeutic antibody that specifically binds the protein having the amino acid sequence of SEQ ID 35 NO:1. The invention further provides an antagonist which specifically binds the protein having the

WO 03/018768

PCT/US02/27144

amino acid sequence of SEQ ID NO:1. The invention yet further provides a small drug molecule which specifically binds the protein having the amino acid sequence of SEQ ID NO:1. The invention also provides a method for testing ligand for effectiveness as an agonist or antagonist comprising exposing a sample comprising the protein to the molecule or compound, and detecting agonist or  
5 antagonist activity in the sample.

The invention provides a method for using a protein to screen a plurality of antibodies to identify an antibody that specifically binds the protein comprising contacting a plurality of antibodies with the protein under conditions to form an antibody:protein complex, and dissociating the antibody from the antibody:protein complex, thereby obtaining antibody that specifically binds the protein. In  
10 one aspect the antibodies are selected from intact immunoglobulin molecule, a polyclonal antibody, a monoclonal antibody, a bispecific molecule, a multispecific molecule, a chimeric antibody, a recombinant antibody, a humanized antibody, single chain antibodies, a Fab fragment, an F(ab)<sub>2</sub> fragment, an Fv fragment, and an antibody-peptide fusion protein. The invention provides purified antibodies which bind specifically to a protein.

The invention also provides methods for using a protein to prepare and purify polyclonal and monoclonal antibodies which specifically bind the protein. The method for preparing a polyclonal antibody comprises immunizing a animal with protein under conditions to elicit an antibody response, isolating animal antibodies, attaching the protein to a substrate, contacting the substrate with isolated antibodies under conditions to allow specific binding to the protein, dissociating the antibodies from  
20 the protein, thereby obtaining purified polyclonal antibodies. The method for preparing a monoclonal antibodies comprises immunizing a animal with a protein under conditions to elicit an antibody response, isolating antibody producing cells from the animal, fusing the antibody producing cells with immortalized cells in culture to form monoclonal antibody producing hybridoma cells, culturing the hybridoma cells, and isolating monoclonal antibodies from culture.

The invention also provides a method for using an antibody to detect expression of a protein  
25 in a sample, the method comprising combining the antibody with a sample under conditions for formation of antibody:protein complexes, and detecting complex formation, wherein complex formation indicates expression of the protein in the sample. In one aspect, the sample is selected from colon or stomach tissue. In a second aspect, complex formation is compared to standards and is  
30 diagnostic of a a colon or stomach cancer.

The invention provides a method for immunopurification of a protein comprising attaching an antibody to a substrate, exposing the antibody to a sample containing protein under conditions to allow antibody:protein complexes to form, dissociating the protein from the complex, and collecting purified protein. The invention also provides a composition comprising an antibody that specifically  
35 binds the protein and a labeling moiety or pharmaceutical agent; a kit comprising the composition; an

WO 03/018768

PCT/US02/27144

array element comprising the antibody; a substrate upon which the antibody is immobilized. The invention further provides a method for using an antibody to assess efficacy of a molecule or compound, the method comprising treating a sample containing protein with a molecule or compound; contacting the protein in the sample with the antibody under conditions for complex formation; determining the amount of complex formation; and comparing the amount of complex formation in the treated sample with the amount of complex formation in an untreated sample, wherein a difference in complex formation indicates efficacy of the molecule or compound.

The invention provides a method for treating colon cancer comprising administering to a subject in need of therapeutic intervention a therapeutic antibody that specifically binds the protein, a bispecific molecule that specifically binds the protein, and a multispecific molecule that specifically binds the protein, or a composition comprising an antibody and a pharmaceutical agent. The invention also provides a method for delivering a pharmaceutical or therapeutic agent to a cell comprising attaching the pharmaceutical or therapeutic agent to a bispecific molecule that specifically binds the protein and administering the bispecific molecule to a subject in need of therapeutic intervention, wherein the bispecific molecule delivers the pharmaceutical or therapeutic agent to the cell. In one aspect, the cell is an epithelial cell of the colon.

The invention provides an agonist that specifically binds the protein, and a composition comprising the agonist and a pharmaceutical carrier. The invention also provides an antagonist that specifically binds the protein, and a composition comprising the antagonist and a pharmaceutical carrier. The invention further provides a pharmaceutical agent or a small drug molecule that specifically binds the protein.

The invention provides an antisense molecule of 18 to 30 nucleotides in length that specifically binds a portion of a polynucleotide having a nucleic acid sequence of SEQ ID NO:2 or the complement thereof wherein the antisense molecule inhibits expression of the protein encoded by the polynucleotide.

The invention also provides an antisense molecule with at least one modified internucleoside linkage or at least one nucleotide analog. The invention further provides that the modified internucleoside linkage is a phosphorothioate linkage and that the modified nucleobase is a 5-methylcytosine.

The invention provides a method for inserting a heterologous marker gene into the genomic DNA of a mammal to disrupt the expression of the endogenous polynucleotide. The invention also provides a method for using a cDNA to produce a mammalian model system, the method comprising constructing a vector containing the cDNA selected from SEQ ID NOs:2-16, transforming the vector into an embryonic stem cell, selecting a transformed embryonic stem cell, microinjecting the transformed embryonic stem cell into a mammalian blastocyst, thereby forming a chimeric blastocyst, transferring the chimeric blastocyst into a pseudopregnant dam, wherein the dam gives birth to a

WO 03/018768

PCT/US02/27144

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 TABLES

Figures 1A through 1H show the transmembrane protein tumor antigen (TMDC; SEQ ID NO:1) encoded by the cDNA (SEQ ID NO:2). The alignment was produced using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA).

Figure 2 shows a hydrophobicity plot for TMDC. The negative Y axis shows hydrophobicity; the X axis, the position/number of the amino acid residue number. The plot was produced using MACDNASIS PRO software.

Figure 3 shows the expression of TMDC in various normal adult tissues. The X-axis indicates the tissue type; the Y-axis, the expression of TMDC relative to that found in normal colon tissue (i.e., set at 100%). QPCR analysis was performed using the TAQMAN protocol (Applied Biosystems (ABI), Foster City CA). Tissues were obtained from Clinomics (Pittsfield MA) and Clontech (Palo Alto CA). The analysis was performed using an oligonucleotide probe extending from about nucleotide 1899 to about nucleotide 1966 of SEQ ID NO:2.

Figure 4 shows the differential expression of TMDC in tissues from patients with colon cancer relative to donor-matched-normal colon tissue using QPCR (ABI). The X-axis lists the patient ID (Donor ID); the Y-axis, the expression TMPTA relative to that found in normal colon tissue (i.e., set at 100%). Tumor samples are displayed in black, and normal tissue in white. The analysis was performed using an oligonucleotide probe extending from about nucleotide 1899 to about nucleotide 1966 of SEQ ID NO:2.

Figure 5 shows the differential expression of TMDC in various colon tumor cell lines compared to a non-tumorigenic colon cell line (LS123) and to that found in normal colon tissue (i.e., set at 100%) using QPCR (ABI). Cell lines were obtained from the ATCC (Manassas VA). The analysis was performed using an oligonucleotide probe extending from about nucleotide 1899 to about nucleotide 1966 of SEQ ID NO:2.

Figure 6 shows the expression of the transcript encoding TMDC in normal colon tissue. Thin sections were stained with DAPI and hybridized *in situ* using sense or antisense RNA probes made from a fragment of SEQ ID NO:2 extending from about nucleotide 1068 to about nucleotide 2324 of SEQ ID NO:2.

Figure 7 shows the expression of the transcript encoding TMDC in a villous adenocarcinoma of the colon. Thin sections were stained with DAPI and hybridized *in situ* using the antisense RNA probe made from a fragment of SEQ ID NO:2 extending from about nucleotide 1068 to about nucleotide 2324 of SEQ ID NO:2.

Table 1 shows the Northern analysis for TMDC produced using the LIFSEQ Gold database

WO 03/018768

PCT/US02/27144

(Incyte Genomics, Palo Alto CA). The first column presents the tissue categories; the second column, the number of clones in the tissue category; the third column, the number of libraries in which at least one transcript was found relative to the total number of libraries in that category; the fourth column, the absolute abundance of the transcript (number of transcripts); and the fifth column, percent abundance of the transcript.

Table 2 shows the Northern analysis for TMDC in tissues of the digestive system in which transcripts are overexpressed, i.e., an abundance >1 transcript is found in any one cDNA library. The first column shows the library identification, the second column, the library description, the third column the absolute abundance (number of transcripts/library), and the fourth column, the percent abundance of the transcript.

Table 3 shows the differential expression of TMDC in tissues from patients with colon cancer relative to normal colon tissue as determined by microarray analysis. The first column lists the differential expression (DE) between the tumor sample and normal tissue. The results are expressed in terms of the ratio of tumor/normal expression. Column 2 (P1 Description) lists the tissue and patient donor (Dn) for microscopically normal samples labeled with the fluorescent green dye, Cy3. Column 3 (P2 Description) lists the tissue and patient donor (Dn) for diseased samples (colon tumor or colon polyps) labeled with the 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" may 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 belongs. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

#### Definitions

"Antibody" refers to intact immunoglobulin molecule, a polyclonal antibody, a monoclonal antibody, a chimeric antibody, a recombinant antibody, a humanized antibody, single chain antibodies, a Fab fragment, an F(ab)<sub>2</sub> fragment, an Fv fragment, and an antibody-peptide fusion protein.

"Antigenic determinant" refers to an antigenic or immunogenic epitope, structural feature, or

WO 03/018768

PCT/US02/27144

region of an oligopeptide, peptide, or protein which is capable of inducing formation of an antibody that specifically binds the protein. Biological activity is not a prerequisite for immunogenicity.

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

A "bispecific molecule" has two different binding specificities and can be bound to two different molecules or two different sites on a molecule concurrently. Similarly, a "multispecific molecule" can bind to multiple (more than two) distinct targets, one of which is a molecule on the surface of an immune cell. Antibodies can perform as or be a part of bispecific or multispecific molecules.

"TMDC" refers to a transmembrane protein that is exactly or highly homologous (>85%) to the amino acid sequence of SEQ ID NO:1 obtained from any species including bovine, ovine, porcine, murine, equine, and preferably the human species, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

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 a nucleic acid molecule under conditions of high stringency.

"cDNA" refers to an isolated polynucleotide, nucleic acid molecule, or any fragment thereof that contains from about 400 to about 12,000 nucleotides. It may have originated recombinantly or synthetically, may be double-stranded or single-stranded, may represent coding and noncoding 3' or 5' sequence, and generally lacks introns.

The phrase "cDNA encoding a protein" refers to a nucleic acid whose sequence closely aligns with sequences that encode conserved regions, motifs or domains 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) and BLAST2 (Altschul *et al.* (1997) *Nucleic Acids Res* 25:3389-3402) which provide identity within the conserved region. Brenner *et al.* (1998; *Proc Natl Acad Sci* 95:6073-6078) who analyzed BLAST for its ability to identify structural homologs by sequence identity found 30% identity is a reliable threshold for sequence alignments of at least 150 residues and 40% is a reasonable threshold for alignments of at least 70 residues (Brenner, page 6076, column 2).

A "composition" refers to the polynucleotide and a labeling moiety; a purified protein and a pharmaceutical carrier or a heterologous, labeling or purification moiety; an antibody and a labeling

WO 03/018768

PCT/US02/27144

moiety or pharmaceutical agent; and the like.

"Derivative" refers to a cDNA or a protein that has been subjected to a chemical modification. Derivatization of a cDNA can involve substitution of a nontraditional base such as queosine or of an analog such as hypoxanthine. These substitutions are well known in the art. Derivatization of a cDNA or a protein can also involve the replacement of a hydrogen by an acetyl, acyl, alkyl, amino, formyl, or morpholino group (for example, 5-methylcytosine). Derivative molecules retain the biological activities of the naturally occurring molecules but may confer longer lifespan or enhanced activity.

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

"Disorder" refers to conditions, diseases or syndromes in which TMDC or the mRNA encoding TMDC are differentially expressed; these include colon and stomach cancer.

An "expression profile" is a representation of gene expression in a sample. A nucleic acid expression profile is produced using sequencing, hybridization, or amplification (quantitative PCR) technologies and mRNAs or cDNAs from a sample. A protein expression profile, although time delayed, mirrors the nucleic acid expression profile and may use antibody or protein arrays, enzyme-linked immunosorbent assays, fluorescence-activated cell sorting, spatial immobilization such as 2D-PAGE, and radioimmunoassays including radiolabeling and quantification using a scintillation counter and western analysis to detect protein expression in a sample. The nucleic acids, proteins, or antibodies may be used in solution or attached to a substrate, and their detection is based on methods and labeling moieties well known in the art. Expression profiles may also be evaluated by methods such as electronic northern analysis, guilt-by-association, and transcript imaging. Expression profiles produced using any of the above methods may be contrasted with expression profiles produced using normal or diseased tissues. Of note is the correspondence between mRNA and protein expression has been discussed by Zweiger (2001, *Transducing the Genome*, McGraw-Hill, San Francisco, CA) and Glavas et al. (2001; T cell activation upregulates cyclic nucleotide phosphodiesterases 8A1 and 7A3, *Proc Natl Acad Sci* 98:6319-6342) among others.

"Fragment" refers to a chain of consecutive nucleotides from about 50 to about 5000 base pairs in length. Fragments may be used in PCR or hybridization technologies to identify related nucleic acid molecules and in binding assays to screen for a ligand. Such ligands are useful as therapeutics to regulate replication, transcription or translation.

"Guilt-by-association" (GBA) is a method for identifying cDNAs or proteins that are associated with a specific disease, regulatory pathway, subcellular compartment, cell type, tissue type, or species by their highly significant co-expression with known markers or therapeutics.

A "hybridization complex" is formed between a cDNA and a nucleic acid of a sample when the

WO 03/018768

PCT/US02/27144

purines of one molecule hydrogen bond with the pyrimidines of the complementary molecule, e.g., 5'-A-G-T-C-3' base pairs with 3'-T-C-A-G-5'. Hybridization conditions, degree of complementarity and the use of nucleotide analogs affect the efficiency and stringency of hybridization reactions.

"Identity" 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), CLUSTALW (Thompson *et al.* (1994) *Nucleic Acids Res* 22:4673-4680), or BLAST2 (Altschul (1997, *supra*). BLAST2 may be used in a standardized and reproducible way to insert gaps in one of the sequences in order to optimize alignment and to achieve a more meaningful comparison between them. "Similarity" uses the same algorithms but takes conservative substitution of residues into account. In proteins, similarity exceeds identity in that substitution of a valine for a leucine or isoleucine, is counted in calculating the reported percentage. Substitutions which are considered to be conservative are well known in the art.

"Isolated" or "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.

"Labeling moiety" refers to any reporter molecule including radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, substrates, cofactors, inhibitors, or magnetic particles than can be attached to or incorporated into a polynucleotide, protein, or antibody. Visible labels and dyes include but are not limited to anthocyanins,  $\beta$  glucuronidase, biotin, BIODIPY, Coomassie blue, Cy3 and Cy5, 4,6-diamidino-2-phenylindole (DAPI), digoxigenin, fluorescein, FITC, gold, green fluorescent protein, lissamine, luciferase, phycoerythrin, rhodamine, spiro red, silver, streptavidin, and the like. Radioactive markers include radioactive forms of hydrogen, iodine, phosphorous, sulfur, and the like.

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

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

A "pharmaceutical agent" may be an antibody, an antisense molecule, a bispecific molecule, a multispecific molecule, a peptide, a protein, a radionuclide, a small drug molecule, a cytospecific or cytotoxic drug such as abrin, actinomycin D, cisplatin, croton, doxorubicin, 5-fluorouracil, methotrexate, ricin, vincristine, vinblastine, or any combination of these elements.

"Post-translational modification" of a protein can involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and the like. These processes may occur

WO 03/018768

PCT/US02/27144

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 determinant of the protein identified using Kyte-Doolittle algorithms of the PROTEAN program (DNASTAR, Madison WI). An "oligopeptide" is an amino acid sequence from about five residues to about 15 residues that is used as part of a fusion protein to produce an antibody.

"Sample" is used in its broadest sense as containing nucleic acids, proteins, and antibodies. A sample may comprise a bodily fluid such as ascites, blood, cerebrospinal fluid, lymph, semen, sputum, urine and the like; 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 biopsy, or a tissue print; buccal cells, skin, hair, a hair follicle; and the like.

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

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

A "transcript image" (TI) is a profile of gene transcription activity in a particular tissue at a particular time. TI provides assessment of the relative abundance of expressed polynucleotides in the cDNA libraries of an EST database as described in USPN 5,840,484, incorporated herein by reference.

"Variant" refers to molecules that are recognized variations of a protein or the polynucleotides that encode it. Splice variants may be determined by BLAST score, wherein the score is at least 100, and most preferably at least 400. Allelic variants have a high percent identity to the cDNAs and may differ by about three bases per hundred bases. "Single nucleotide polymorphism" (SNP) refers to a change in a single base as a result of a substitution, insertion or deletion. The change may be conservative (purine for purine) or non-conservative (purine to pyrimidine) and may or may not result in a change in an encoded

WO 03/018768

PCT/US02/27144

amino acid or its secondary, tertiary, or quaternary structure.

#### THE INVENTION

The invention is based on the discovery of a transmembrane protein differentially expressed in cancer, a cDNA which encodes the protein and an antibody that specifically binds the protein. The protein, or portions thereof, the cDNA, or fragments thereof, and the antibody can be used directly or as compositions to diagnose, to stage, to treat, or to monitor the progression or treatment of colon or stomach cancer.

Nucleic acids encoding the TMDC of the present invention were first identified in Incyte Clone 1929823 from a colon tumor library (COLNTUTO3) using a computer search for nucleotide and/or amino acid sequence alignments. SEQ ID NO:2 was derived from the following overlapping and/or extended nucleic acid sequences (SEQ ID NO:3-10) and their associated cDNA libraries: Incyte Clones 1929823H1, 1929823T6, and 1341151F6 (COLNTUTO3), 7703595H1 (UTRETUE01), 8146316H1 (MIXDTME01), 3274531H1, shotgun sequences SCCA02331V1 and SCCA04417V, and genomic sequence g2951946\_010 (SEQ ID NO:11).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:1 as shown in Figures 1A through 1H. TMDC is 760 amino acids in length and has seven potential N-glycosylation sites at amino acid residues N16, N77, N221, N264, N342, N350, and N567. TMDC has one potential cyclic-AMP/cyclic-GMP-dependent protein kinase phosphorylation site at T237, eleven potential casein kinase phosphorylation sites at S46, S48, S75, S97, T115, T129, S174, T241, S474, S634, and S752, six potential protein kinase C phosphorylation sites at S59, T115, T148, T188, S640, and S749, and one potential tyrosine kinase phosphorylation site at Y536. HMMR analysis indicates the presence of nine transmembrane domains as follows: TM-1, amino acid residues 210-230; TM-2, amino acid residues 281-299; TM-3, amino acid residues 372-392; TM-4, amino acid residues 447-467; TM-5, amino acid residues 487-507; TM-6, amino acid residues 540-562; TM-7, amino acid residues 586-610; TM-8, amino acid residues 654-672; and TM-9, amino acid residues 956-974. Useful antigenic epitopes of SEQ ID NO:1 extend from about amino acid residue S110 to about R150, from about F230 to about G270, from about V330 to about F370, and from about C420 to about I450. An antibody which specifically binds transmembrane protein tumor antigen is useful in a diagnostic assay to identify a cancer, in particular colon or stomach cancer.

Figure 2 is a hydrophobicity plot for TMDC that shows the various transmembrane regions as hydrophobic regions (negative values on the Y axis of the plot).

Figure 3 shows the results of various normal adult tissues analyzed for TMDC expression by TAQMAN analysis. The most significant expression of TMDC, was found in testis, adipose tissue, breast duodenum, and colon, indicating that TMDC has a relatively restricted normal tissue distribution. The high expression in testis, however, was associated with a higher than normal expression in the internal

WO 03/018768

PCT/US02/27144

control,  $\beta 2$  microglobulin.

Table 1 shows the expression of the TMDC across tissue categories by northern analysis of cDNA libraries in the LIFESEQ Gold database (Incye Genomics). The results show the highest abundance (total number of transcripts found) of TMDC in digestive system. The differences observed between the results of Table 1 and Figure 3, above, most likely reflect the high incidence of fetal and diseased tissues in cDNA libraries of the LIFESEQ database.

Table 2 further shows that within the digestive system the cDNA libraries overexpressing TMDC (>1 transcript/library) are diseased tissues including colon tumors, FAP, inflamed intestine, and stomach tumor. Particularly noteworthy is the overexpression of TMDC in a colon tumor (COLNTUT03) matched with normal colon tissue from the same donor (COLNNOT16) in which TMDC expression was undetectable, and in the stomach tumor, STOMTUP02, which showed the highest abundance of any digestive tissue expressing TMDC.

Figure 4 shows the expression of TMDC in colon cancer tissue samples compared with normal colon tissue using QPCR analysis (Applied Biosystems). The results show increased expression of TMDC in colon tumors in eight of nine samples examined. The results were considered significant if at least a 1.2-fold difference in expression was observed between cancerous and normal tissue.

Figure 5 similarly shows the expression of TMDC in various human colon tumor cell lines compared to a non-tumorigenic colon cell line, LS123, using QPCR analysis. TMDC is overexpressed in six of eight colon tumor cell lines examined, i.e., LS174, HCT116, Caco2, HT29, COLO205, and SW620.

Table 3 shows the results of microarray analysis comparing the expression of TMDC in colon cancer tissues relative to normal colon tissue. The results show an increased expression of TMDC in two of 14 patients examined. Differential expression (column 1) was considered significant if at least a 1.5-fold difference in expression was observed between cancerous and normal tissue. Differences in relative expression values for samples analyzed by QPCR in Figure 3 compared to Table 1 is likely due, in part, to the greater sensitivity and larger dynamic range for QPCR analysis than for microarray analysis.

Mammalian variants of the cDNA encoding TMDC were identified using BLAST2 with default parameters and the ZOOSEQ databases (Incye Genomics). These preferred variants have from about 84% to 90% identity as shown in the table below. The first column represents the SEQ IDvar for variant cDNAs; the second column, the clone number for the variant cDNAs; the third column, the species; the fourth column, the percent identity to the human cDNA; and the fifth column, the alignment of the variant cDNA to the human cDNA.

SEQ ID <sub>var</sub>	cDNA <sub>var</sub>	Species	Identity	N <sub>t</sub> <sub>1</sub> Alignment
12	701294553H1	Rat	85%	474-654
13	701600294H1	Rat	88%	2927-2994
14	2016808H1	Mouse	89%	2939-3052
15	239780_Mm.1	Mouse	84%	714-1615
16	703528478I1	Dog	90%	2927-2990

WO 03/018768

PCT/US02/27144

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

The cDNAs of SEQ ID NOs:2-16 may be used in hybridization, amplification, and screening technologies to identify and distinguish among SEQ ID NO:2 and related molecules in a sample. The mammalian cDNAs, SEQ ID NOs:12-16, may be used to produce transgenic cell lines or organisms which are model systems for human a colon or stomach cancer and upon which the toxicity and efficacy of therapeutic treatments may be tested. Toxicology studies, clinical trials, and subject/patient treatment profiles may be performed and monitored using the cDNAs, proteins, antibodies and molecules and compounds identified using the cDNAs and proteins of the present invention.

#### 15 **Characterization and Use of the Invention**

##### cDNA libraries

In a particular embodiment disclosed herein, mRNA is isolated from mammalian cells and tissues using methods which are well known to those skilled in the art and used to prepare the cDNA libraries.

The Incyte cDNAs were isolated from mammalian cDNA libraries prepared as described in the

20 **EXAMPLES.** The consensus sequence is present in a single clone insert, or chemically assembled, based on the electronic assembly from sequenced fragments including Incyte cDNAs and extension and/or shotgun sequences. Computer programs, such as PHRAP (P Green, University of Washington, Seattle WA) and the AUTOASSEMBLER application (ABI), are used in sequence assembly and are described in **EXAMPLE V.** After verification of the 5' and 3' sequence, Incyte clone 1929823F6 which encodes TMDC was designated a reagent for research and development.

##### 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 Biosciences (APB), Piscataway NJ), or combinations of polymerases and proofreading exonucleases (Invitrogen, Carlsbad CA). 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) and sequencing, with the PRISM 3700, 377 or 373 DNA sequencing systems (ABI) or the MEGABACE 1000 DNA sequencing system (APB).

35 The nucleic acid sequences of the cDNAs presented in the Sequence Listing were prepared by

WO 03/018768

PCT/US02/27144

such automated methods and may contain occasional sequencing errors and unidentified nucleotides, designated with an N, that reflect state-of-the-art technology at the time the cDNA was sequenced. Vector, linker, and polyA sequences were masked using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. Ns and SNPs can be verified either by resequencing the cDNA or using algorithms to compare multiple sequences that overlap the area in which the Ns or SNP occur. Both of these techniques are well known to and used by those skilled in the art. The sequences may be analyzed using a variety of algorithms described in Ausubel et al. (1997; Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7) and in Meyers (1995; Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

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

#### 20 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 (ABI), nested primers, and cDNA or genomic DNA libraries may be used to extend the nucleic acid sequence. For all PCR-based methods, primers may be designed using 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, genomic, rather than cDNA libraries are used.

#### 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 TMDC, allelic variants, or related molecules. The probe may be DNA or RNA, may be single-stranded, and should have at least 50% sequence identity to any of the nucleic acid sequences, SEQ ID NOs:2-9. Hybridization probes may be produced using oligolabeling, nick-translation, end-labeling, or PCR amplification in the

WO 03/018768

PCT/US02/27144

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 kits such as those provided by APB.

The stringency of hybridization is determined by G+C content of the probe, salt concentration, and temperature. In particular, stringency can be increased by reducing the concentration of salt or raising the hybridization temperature. Hybridization can be performed at low stringency with buffers, such as 5xSSC with 1% sodium dodecyl sulfate (SDS) at 60C, which permits the formation of a hybridization complex between nucleic acid sequences that contain some mismatches. Subsequent washes are performed at higher stringency with buffers such as 0.2xSSC with 0.1% SDS at either 45C (medium stringency) or 68C (high stringency). At high stringency, hybridization complexes will remain stable only where the nucleic acids are completely complementary. In some membrane-based hybridizations, from about 35% to about 50% formamide can be added to the hybridization solution to reduce the temperature at which hybridization is performed. Background signals can be reduced by the use of detergents such as Sarkosyl or TRITON X-100 (Sigma-Aldrich) and a blocking agent such as denatured salmon sperm DNA.

15 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 well known in the art. Oligonucleotides or cDNAs may be used as hybridization probes or targets to monitor the expression level of large numbers of genes simultaneously or to identify genetic variants, mutations, and single nucleotide polymorphisms.

Arrays may be used to determine gene function; to understand the genetic basis of a condition, disease, or disorder; to diagnose a condition, disease, or disorder; and to develop and monitor the activities of therapeutic agents. (See, e.g., USPN 5,474,796; Schena *et al.* (1996) *Proc Natl Acad Sci* 93:10614-10619; Heller *et al.* (1997) *Proc Natl Acad Sci* 94:2150-2155; USPN 5,605,662.)

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

### 30 QPCR

QPCR is a method for quantifying a nucleic acid molecule based on detection of a fluorescent signal produced during PCR amplification (Gibson *et al.* (1996) *Genome Res* 6:995-1001; Heid *et al.* (1996) *Genome Res* 6:986-994). Amplification is carried out on machines such as the PRISM 7700 detection system (ABI) which consists of a 96-well thermal cycler connected to a laser and charge-coupled device (CCD) optics system. To perform QPCR, a PCR reaction is carried out in the presence of a doubly

35

WO 03/018768

PCT/US02/27144

labeled probe. The probe, which is designed to anneal between the standard forward and reverse PCR primers, is labeled at the 5' end by a fluorescent reporter dye such as 6-carboxyfluorescein (6-FAM) and at the 3' end by a quencher molecule such as 6-carboxy-tetramethyl-rhodamine (TAMRA). As long as the probe is intact, the 3' quencher extinguishes fluorescence by the 5' reporter. However, during each primer extension cycle, the annealed probe is degraded as a result of the intrinsic 5' to 3' nuclease activity of Taq polymerase (Holland *et al.* (1991) Proc Natl Acad Sci 88:7276-7280). This degradation separates the reporter from the quencher, and fluorescence is detected every few seconds by the CCD. The higher the starting copy number of the nucleic acid, the sooner an increase in fluorescence is observed. A cycle threshold ( $C_T$ ) value, representing the cycle number at which the PCR product crosses a fixed threshold of detection is determined by the instrument software. The  $C_T$  is inversely proportional to the copy number of the template and can therefore be used to calculate either the relative or absolute initial concentration of the nucleic acid molecule in the sample. The relative concentration of two different molecules can be calculated by determining their respective  $C_T$  values (comparative  $C_T$  method). Alternatively, the absolute concentration of the nucleic acid molecule can be calculated by constructing a standard curve using a housekeeping molecule of known concentration. The process of calculating  $C_T$  values, preparing a standard curve, and determining starting copy number is performed using SEQUENCE DETECTOR 1.7 software (ABI).

#### Expression

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

A variety of host systems may be transformed with an expression vector. These include, 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 or plant cell systems transformed with expression vectors containing viral and/or bacterial elements (Ausubel *supra*, unit 16). In mammalian cell systems, an adenovirus transcriptional/translational complex may be utilized. 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

WO 03/018768

PCT/US02/27144

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 (Invitrogen).

5 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 10 transformation, cells are allowed to grow for about 1 to 2 days in enriched media and then are transferred to selective media. Selectable markers, antimetabolite, antibiotic, or herbicide resistance genes, confer resistance to the relevant selective agent and allow growth and recovery of cells which successfully express the introduced sequences. Resistant clones identified either by survival on selective media or by the expression of visible markers may be propagated using culture techniques. Visible markers are also 15 used to estimate the amount of protein expressed by the introduced genes. Verification that the host cell contains the desired cDNA is based on DNA-DNA or DNA-RNA hybridizations or PCR amplification.

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 20 specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities may be chosen to ensure the correct modification and processing of the recombinant protein.

#### Recovery of Proteins from Cell Culture

Heterologous moieties engineered into a vector for ease of purification include glutathione S- 25 transferase (GST), 6xHis, FLAG, MYC, and the like. GST and 6-His are purified using affinity matrices such as immobilized glutathione and metal-chelate resins, respectively. FLAG and MYC are purified using 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 30 Ausubel (supra, unit 16).

#### Protein Identification

Several techniques have been developed which permit rapid identification of proteins using high performance liquid chromatography and mass spectrometry (MS). Beginning with a sample containing proteins, the method is: 1) proteins are separated using two-dimensional gel electrophoresis (2-DE), 2) 35 selected proteins are excised from the gel and digested with a protease to produce a set of peptides; and 3)

WO 03/018768

PCT/US02/27144

the peptides are subjected to mass spectral analysis to derive peptide ion mass and spectral pattern information. The MS information is used to identify the protein by comparing it with information in a protein database (Shevenko *et al.* (1996) *Proc Natl Acad Sci* 93:14440-14445). Proteins are separated by 2DE employing isoelectric focusing (IEF) in the first dimension followed by 5 SDS-PAGE in the second dimension. For IEF, an immobilized pH gradient strip is useful to increase reproducibility and resolution of the separation. Alternative techniques may be used to improve resolution of very basic, hydrophobic, or high molecular weight proteins. The separated proteins are detected using a stain or dye such as silver stain, Coomassie blue, or spyro red (Molecular Probes, Eugene OR) that is compatible with MS. Gels may be blotted onto a PVDF membrane for western analysis and optically 10 scanned using a STORM scanner (APB) to produce a computer-readable output which is analyzed by pattern recognition software such as MELANIE (GeneBio, Geneva, Switzerland). The software annotates individual spots by assigning a unique identifier and calculating their respective x,y coordinates, molecular masses, isoelectric points, and signal intensity. Individual spots of interest, such as those representing differentially expressed proteins, are excised and proteolytically digested with a site-specific protease such 15 as trypsin or chymotrypsin, singly or in combination, to generate a set of small peptides, preferably in the range of 1-2 kDa. Prior to digestion, samples may be treated with reducing and alkylating agents, and following digestion, the peptides are then separated by liquid chromatography or capillary electrophoresis and analyzed using MS.

MS converts components of a sample into gaseous ions, separates the ions based on their 20 mass-to-charge ratio, and determines relative abundance. For peptide mass fingerprinting analysis, a MALDI-TOF (Matrix Assisted Laser Desorption/Ionization-Time of Flight), ESI (Electrospray Ionization), and TOF-TOF (Time of Flight/Time of Flight) machines are used to determine a set of highly accurate peptide masses. Using analytical programs, such as TURBOSEQUENT software (Finnigan, San Jose CA), the MS data is compared against a database of theoretical MS data derived from known or 25 predicted proteins. A minimum match of three peptide masses is used for reliable protein identification. If additional information is needed for identification, Tandem-MS may be used to derive information about individual peptides. In tandem-MS, a first stage of MS is performed to determine individual peptide masses. Then selected peptide ions are subjected to fragmentation using a technique such as collision induced dissociation (CID) to produce an ion series. The resulting fragmentation ions are analyzed in a 30 second round of MS, and their spectral pattern may be used to determine a short stretch of amino acid sequence (Dancik *et al.* (1999) *J Comput Biol* 6:327-342).

Assuming the protein is represented in the database, a combination of peptide mass and fragmentation data, together with the calculated MW and pI of the protein, will usually yield an unambiguous identification. If no match is found, protein sequence can be obtained using direct chemical sequencing 35 procedures well known in the art (cf. Creighton (1984) Proteins, Structures and Molecular Properties. WH

WO 03/018768

PCT/US02/27144

Freeman, New York NY).

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 431A peptide synthesizer (ABI). A protein or portion thereof may be purified by preparative high performance liquid chromatography and its composition confirmed by amino acid analysis or by sequencing (Creighton (1984) Proteins. Structures and Molecular Properties, WH Freeman, New York NY).

Antibodies

Antibodies, or immunoglobulins (Ig), are components of immune response expressed on the surface of or secreted into the circulation by B cells. The prototypical antibody is a tetramer composed of two identical heavy polypeptide chains (H-chains) and two identical light polypeptide chains (L-chains) interlinked by disulfide bonds which binds and neutralizes foreign antigens. Based on their H-chain, antibodies are classified as IgA, IgD, IgE, IgG or IgM. The most common class, IgG, is tetrameric while other classes are variants or multimers of the basic structure.

Antibodies are described in terms of their two functional domains. Antigen recognition is mediated by the Fab (antigen binding fragment) region of the antibody, while effector functions are mediated by the Fc (crystallizable fragment) region. The binding of antibody to antigen triggers destruction of the antigen by phagocytic white blood cells such as macrophages and neutrophils. These cells express surface Fc receptors that specifically bind to the Fc region of the antibody and allow the phagocytic cells to destroy antibody-bound antigen. Fc receptors are single-pass transmembrane glycoproteins containing about 350 amino acids whose extracellular portion typically contains two or three Ig domains (Sears *et al.* (1990) *J Immunol* 144:371-378).

WO 03/018768

PCT/US02/27144

Preparation and Screening of Antibodies

Various hosts including mice, rats, rabbits, goats, llamas, camels, and human cell lines may be immunized by injection with an antigenic determinant. Adjuvants such as Freund's, mineral gels, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, 5 keyhole limpet hemacyanin (KLH; Sigma-Aldrich), and dinitrophenol may be used to increase immunological response. In humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum increase response. The antigenic determinant may be an oligopeptide, peptide, or protein. When the amount of antigenic determinant allows immunization to be repeated, specific polyclonal antibody with high affinity can be obtained (Klinman and Press (1975) Transplant Rev 24:41-83). Oligopeptides which 10 may contain between about five and about fifteen amino acids identical to a portion of the endogenous protein 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 the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler *et al.* (1975) Nature 256:495-497; 15 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).

Chimeric antibodies may be produced by techniques such as splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity (Morrison *et al.* (1984) Proc Natl Acad Sci 81:6851-6855; Neuberger *et al.* (1984) Nature 312:604-608; 20 and Takeda *et al.* (1985) Nature 314:452-454). Alternatively, techniques described for antibody production may be adapted, using methods known in the art, to produce specific, single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton (1991) Proc Natl Acad Sci 88:10134-10137). Antibody fragments which contain specific binding sites for an antigenic determinant 25 may also be produced. For example, such fragments include, but are not limited to, F(ab)<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab)<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse *et al.* (1989) Science 246:1275-1281).

Antibodies may also be produced by inducing production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in Orlandi *et al.* (1989; Proc Natl Acad Sci 86:3833-3837) or Winter *et al.* (1991; Nature 349:293-299). A protein may be used in screening assays of phagemid or B-lymphocyte immunoglobulin libraries to identify antibodies having a desired specificity. Numerous protocols for competitive binding or immunoassays using either 35 polyclonal or monoclonal antibodies with established specificities are well known in the art.

WO 03/018768

PCT/US02/27144

#### Antibody Specificity

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

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing about 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of protein-antibody complexes. Procedures for making antibodies, evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are discussed in Catty (*supra*) and Ausubel (*supra*) pp. 11.1-11.31.

#### Cell Transformation Assays

Cell transformation, the conversion of a normal cell to a cancerous cell, is a highly complex and genetically diverse process. However, certain alterations in cell physiology that are associated with this process can be assayed using either *in vitro* cell-based systems or *in vivo* animal models. Known alterations include acquired self-sufficiency relative to growth signals, an insensitivity to growth-inhibitory signals, unlimited replicative potential, evasion of apoptosis, sustained angiogenesis, and cellular invasion and metastasis. (See Hanahan and Weinberg (2000) Cell 100:57-70.) Such assays can be used, for example, to assess the effect of overexpression of a gene such as TMDC in a cell, on cell transformation.

#### DIAGNOSTICS

Differential expression of TMDC, as detected using TMDC, cDNA encoding TMDC, or an antibody that specifically binds TMDC, and at least one of the assays below can be used to diagnose a colon or stomach cancer.

#### Labeling of Molecules for Assay

WO 03/018768

PCT/US02/27144

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 kits such as those supplied by Promega (Madison WI) or APB for incorporation of a labeled nucleotide such as <sup>32</sup>P-dCTP (APB), Cy3-dCTP or Cy5-dCTP (Qiagen-Operon, 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).

#### Nucleic Acid Assays

The cDNAs, fragments, oligonucleotides, complementary RNAs, and peptide nucleic acids (PNA) may be used to detect and quantify differential gene expression for diagnosis of a disorder. Similarly antibodies which specifically bind TMDC may be used to quantitate the protein. Disorders associated with such differential expression include a colon or stomach cancer. The diagnostic assay may use hybridization or amplification technology to compare gene expression in a biological sample from a patient to standard samples in order to detect differential gene expression. Qualitative or quantitative methods for this comparison are well known in the art.

#### Expression Profiles

An expression profile comprises the expression of a plurality of cDNAs or protein as measured using standard assays with a sample. The cDNAs, proteins or antibodies of the invention may be used as elements on a array to produce an expression profile. In one embodiment, the array is used to diagnose or monitor the progression of disease.

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 altered in comparison to either a normal or disease standard, then differential expression indicates the presence of a disorder.

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

WO 03/018768

PCT/US02/27144

By analyzing changes in patterns of gene expression, disease can be diagnosed at earlier stages before the patient is symptomatic. The invention can be used to formulate a prognosis and to design a treatment regimen. The invention can also be used to monitor the efficacy of treatment. For treatments with known side effects, the array is employed to improve the treatment regimen. A dosage is established  
5 that causes a change in genetic expression patterns indicative of successful treatment. Expression patterns associated with the onset of undesirable side effects are avoided. This approach may be more sensitive and rapid than waiting for the patient to show inadequate improvement, or to manifest side effects, before altering the course of treatment.

In another embodiment, animal models which mimic a human disease can be used to characterize  
10 expression profiles associated with a particular condition, disease, or disorder; or treatment of the condition, disease, or disorder. Novel treatment regimens may be tested in these animal models using arrays to establish and then follow expression profiles over time. In addition, arrays may be used with cell cultures or tissues removed from animal models to rapidly screen large numbers of candidate drug molecules, looking for ones that produce an expression profile similar to those of known therapeutic  
15 drugs, with the expectation that molecules with the same expression profile will likely have similar therapeutic effects. Thus, the invention provides the means to rapidly determine the molecular mode of action of a drug.

Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies or in clinical trials or to monitor the treatment of an individual patient. Once the  
20 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 years.

#### Protein Assays

25 Immunological methods for detecting and measuring complex formation as a measure of protein expression using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include antibody arrays, enzyme-linked immunosorbent assays, fluorescence-activated cell sorting, 2D-PAGE and scintillation counting, protein arrays, radioimmunoassays, and western analysis. Such immunoassays typically involve the measurement of complex formation between the  
30 protein and its specific antibody. These assays and their quantitation against purified, labeled standards are well known in the art (Ausubel, supra, unit 10.1-10.6). A two-site, monoclonal-based immunoassay utilizing antibodies reactive to two non-interfering epitopes is preferred, but a competitive binding assay may be employed (Pound (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

These methods are also useful for diagnosing diseases that show differential protein expression.  
35 Normal or standard values for protein expression are established by combining body fluids or cell extracts

WO 03/018768

PCT/US02/27144

taken from a normal mammalian or human subject with specific antibodies to a protein under conditions for complex formation. Standard values for complex formation in normal and diseased tissues are established by various methods, often photometric means. Then complex formation as it is expressed in a subject sample is compared with the standard values. Deviation from the normal standard and toward the diseased standard provides parameters for disease diagnosis or prognosis while deviation away from the diseased and toward the normal standard may be used to evaluate treatment efficacy.

Recently, antibody arrays have allowed the development of techniques for high-throughput screening of recombinant antibodies. Such methods use robots to pick and grid bacteria containing antibody genes, and a filter-based ELISA to screen and identify clones that express antibody fragments. Because liquid handling is eliminated and the clones are arrayed from master stocks, the same antibodies can be spotted multiple times and screened against multiple antigens simultaneously. Antibody arrays are highly useful in the identification of differentially expressed proteins. (See de Wildt *et al.* (2000) *Nature Biotechnol* 18:989-94.)

#### THERAPEUTICS

Chemical and structural similarity, in particular the transmembrane domains, exists between regions of TMDC (SEQ ID NO:1) and other transmembrane proteins. In addition, differential expression is highly associated with colon and stomach cancer. TMDC clearly plays a role in a colon or stomach cancer.

In one embodiment, when decreased expression or activity of the protein is desired, an antibody, antagonist, inhibitor, a pharmaceutical agent or a composition containing one or more of these molecules may be delivered to a subject in need of such treatment. Such delivery may be effected by methods well known in the art and may include delivery by an antibody that specifically binds the protein. For therapeutic use, monoclonal antibodies are used to block an active site, inhibit dimer formation, trigger apoptosis and the like.

In another embodiment, when increased expression or activity of the protein is desired, the protein, an agonist, an enhancer, a pharmaceutical agent or a composition containing one or more of these molecules may be delivered to a subject in need of such treatment. Such delivery may be effected by methods well known in the art and may include delivery of a pharmaceutical agent by an antibody specifically targeted to the protein.

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

WO 03/018768

PCT/US02/27144

Modification of Gene Expression Using Nucleic Acids

Gene expression may be modified by designing complementary or antisense molecules (DNA, RNA, or PNA) to the control, 5', 3', or other regulatory regions of the gene encoding TMDC. Oligonucleotides designed to inhibit transcription initiation are preferred. Similarly, inhibition can be achieved using triple helix base-pairing which inhibits the binding of polymerases, transcription factors, or regulatory molecules (Gee et al. In: Huber and Carr (1994) Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary molecule may also be designed to block translation by preventing binding between ribosomes and mRNA. In one alternative, a library or plurality of cDNAs may be screened to identify those which specifically bind a regulatory, nontranslated sequence.

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

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

cDNA Therapeutics

The cDNAs of the invention can be used in gene therapy. cDNAs can be delivered *ex vivo* to target cells, such as cells of bone marrow. Once stable integration and transcription and or translation are confirmed, the bone marrow may be reintroduced into the subject. Expression of the protein encoded by the cDNA may correct a disorder associated with mutation of a normal sequence, reduction or loss of an endogenous target protein, or overexpression of an endogenous or mutant protein. Alternatively, cDNAs may be delivered *in vivo* using vectors such as retrovirus, adenovirus, adeno-associated virus, herpes simplex virus, and bacterial plasmids. Non-viral methods of gene delivery include cationic liposomes, polylysine conjugates, artificial viral envelopes, and direct injection of DNA (Anderson (1998) Nature

WO 03/018768

PCT/US02/27144

392:25-30; Dachs *et al.* (1997) *Oncol Res* 9:313-325; Chu *et al.* (1998) *J Mol Med* 76(3-4):184-192; Weiss *et al.* (1999) *Cell Mol Life Sci* 55(3):334-358; Agrawal (1996) *Antisense Therapeutics*, Humana Press, Totowa NJ; and August *et al.* (1997) *Gene Therapy* (Advances in Pharmacology, Vol. 40), Academic Press, San Diego CA).

#### 5 Monoclonal Antibody Therapeutics

Antibodies, and in particular monoclonal antibodies, that specifically bind a particular protein, enzyme, or receptor and block its overexpression are now being used therapeutically. The first widely accepted therapeutic antibodies were HERCEPTIN (Trastuzumab, Genentech, S. San Francisco CA) and GLEEVEEC (imatinib mesylate, Novartis Pharmaceuticals, East Hanover NJ). HERCEPTIN is a  
10 humanized antibody approved for the treatment of HER2 positive metastatic breast cancer. It is designed to bind and block the function of overexpressed HER2 protein. GLEEVEEC is indicated for the treatment of patients with Philadelphia chromosome positive (Ph+) chronic myeloid leukemia (CML) in blast crisis, accelerated phase, or in chronic phase after failure of interferon-alpha therapy. A second indication for GLEEVEEC is treatment of patients with KIT (CD117) positive unresectable and/or metastatic malignant  
15 gastrointestinal stromal tumors. Other monoclonal antibodies are in various stages of clinical trials for indications such as prostate cancer, lymphoma, melanoma, pneumococcal infections, rheumatoid arthritis, psoriasis, systemic lupus erythematosus, and the like.

#### Screening and Purification Assays

The cDNA encoding TMDC may be used to screen a library or a plurality of molecules or  
20 compounds for specific binding affinity. The libraries may be antisense molecules, artificial chromosome constructions, branched nucleic acid molecules, DNA molecules, peptides, peptide nucleic acid, proteins such as transcription factors, enhancers, or repressors, RNA molecules, ribozymes, and other ligands which regulate the activity, replication, transcription, or translation of the endogenous gene. The assay involves combining a polynucleotide with a library or plurality of molecules or compounds under  
25 conditions allowing specific binding, and detecting specific binding to identify at least one molecule which specifically binds the cDNA.

In one embodiment, the cDNA of the invention may be incubated with a plurality of purified molecules or compounds and binding activity determined by methods well known in the art, e.g., a gel-retardation assay (USPN 6,010,849) or a reticulocyte lysate transcriptional assay. In another embodiment,  
30 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.

35 In another embodiment, the cDNA may be used to purify a molecule or compound using affinity

WO 03/018768

PCT/US02/27144

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.

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

10 In a preferred embodiment, TMDC 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  
15 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 can be measured. Depending on the particular kind of molecules or compounds being screened, the assay may be used to identify agonists, antagonists, antibodies, DNA molecules, small drug molecules, immunoglobulins, inhibitors, mimetics, peptides, peptide nucleic acids, proteins, and RNA molecules or any other ligand,  
20 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  
25 drug screening assays in which neutralizing antibodies capable of binding the protein specifically compete with a test compound capable of binding to the protein. Molecules or compounds identified by screening may be used in a mammalian model system to evaluate their toxicity or therapeutic potential.

#### Pharmaceutical Compositions

Pharmaceutical compositions may be formulated and administered, to a subject in need of such  
30 treatment, to attain a therapeutic effect. Such compositions contain the instant protein, agonists, antagonists, bispecific molecules, small drug molecules, immunoglobulins, inhibitors, mimetics, multispecific molecules, peptides, peptide nucleic acids, pharmaceutical agent, proteins, and RNA molecules. Compositions may be manufactured by conventional means such as mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing. The  
35 composition may be provided as a salt, formed with acids such as hydrochloric, sulfuric, acetic, lactic,

WO 03/018768

PCT/US02/27144

tartaric, malic, and succinic, or as a lyophilized powder which may be combined with a sterile buffer such as saline, dextrose, or water. These compositions may include auxiliaries or excipients which facilitate processing of the active compounds.

Auxiliaries and excipients may include coatings, fillers or binders including sugars such as lactose, sucrose, mannitol, glycerol, or sorbitol; starches from corn, wheat, rice, or potato; proteins such as albumin, gelatin and collagen; cellulose in the form of hydroxypropylmethyl-cellulose, methyl cellulose, or sodium carboxymethylcellulose; gums including arabic and tragacanth; lubricants such as magnesium stearate or talc; disintegrating or solubilizing agents such as the, agar, alginate, sodium alginate or cross-linked polyvinyl pyrrolidone; stabilizers such as carbopol gel, polyethylene glycol, or titanium dioxide; and dyestuffs or pigments added for identify the product or to characterize the quantity of active compound or dosage.

These compositions may be administered by any number of routes including oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal.

The route of administration and dosage will determine formulation; for example, oral administration may be accomplished using tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, or suspensions; parenteral administration may be formulated in aqueous, physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Suspensions for injection may be aqueous, containing viscous additives such as sodium carboxymethyl cellulose or dextran to increase the viscosity, or oily, containing lipophilic solvents such as sesame oil or synthetic fatty acid esters such as ethyl oleate or triglycerides, or liposomes. Penetrants well known in the art are used for topical or nasal administration.

#### Toxicity and Therapeutic Efficacy

A therapeutically effective dose refers to the amount of active ingredient which ameliorates symptoms or condition. For any compound, a therapeutically effective dose can be estimated from cell culture assays using normal and neoplastic cells or in animal models. Therapeutic efficacy, toxicity, concentration range, and route of administration may be determined by standard pharmaceutical procedures using experimental animals.

The therapeutic index is the dose ratio between therapeutic and toxic effects—LD50 (the dose lethal to 50% of the population)/ED50 (the dose therapeutically effective in 50% of the population)—and large therapeutic indices are preferred. Dosage is within a range of circulating concentrations, includes an ED50 with little or no toxicity, and varies depending upon the composition, method of delivery, sensitivity of the patient, and route of administration. Exact dosage will be determined by the practitioner in light of factors related to the subject in need of the treatment.

Dosage and administration are adjusted to provide active moiety that maintains therapeutic effect.

WO 03/018768

PCT/US02/27144

Factors for adjustment include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular composition.

Normal dosage amounts may vary from 0.1  $\mu\text{g}$ , up to a total dose of about 1 g, depending upon the route of administration. The dosage of a particular composition may be lower when administered to a patient in combination with other agents, drugs, or hormones. Guidance as to particular dosages and methods of delivery is provided in the pharmaceutical literature. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Mack Publishing, Easton PA).

#### 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, gestation period, numbers of progeny, and abundant reference literature. Inbred and outbred rodent strains provide a convenient model for investigation of the physiological consequences of under- or over-expression of genes of interest and for the development of methods for diagnosis and treatment of diseases. A mammal inbred to over-express a particular gene (for example, secreted in milk) may also serve as a convenient source of the protein expressed by that gene.

#### Toxicology

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

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

WO 03/018768

PCT/US02/27144

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.

5 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 test whether long term administration may elicit toxicity, teratogenesis, or carcinogenesis. When studies are conducted on rats, a 10 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 15 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

20 Embryonic (ES) stem cells isolated from rodent embryos retain the ability to form embryonic tissues. When ES cells are placed inside a carrier embryo, they resume normal development and contribute to tissues of the live-born animal. ES cells are the preferred cells used in the creation of experimental knockout and knockin rodent strains. Mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and are grown under culture conditions well known in the art. 25 Vectors used to produce a transgenic strain contain a disease gene candidate and a marker gene, the latter serves to identify the presence of the introduced disease gene. The vector is transformed into ES cells by methods well known in the art, and transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous 30 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.

#### 35 Knockout Analysis

WO 03/018768

PCT/US02/27144

In gene knockout analysis, a region of a 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 pharmaceutical agents to obtain information on treatment of the analogous human condition. These methods have been used to model several human diseases.

#### Non-Human Primate Model

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

In additional embodiments, the cDNAs which encode the 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**

##### **I cDNA Library Construction**

The COLNTUTO3 library was constructed using RNA isolated from colon tumor tissue removed from the sigmoid colon of a 62-year-old Caucasian male during a sigmoidectomy and permanent

WO 03/018768

PCT/US02/27144

colostomy. Pathology indicated grade 2 adenocarcinoma with invasion through the muscularis.

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

The mRNA was handled according to the recommended protocols in the SUPERSERIPPT 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 (Incye Genomics). The plasmid pINCY was subsequently transformed into DH5 $\alpha$  competent cells (Life Technologies).

## II Isolation, Preparation, and Sequencing of cDNAs

Plasmids were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or REAL PREP 96 plasmid purification kit from Qiagen. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao (1994) Anal Biochem 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the CATALYST 800 (ABI) thermal cycler or the DNA ENGINE thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents obtained from APB or supplied in sequencing kits such as the PRISM BIGDYE Terminator cycle sequencing ready reaction kit

WO 03/018768

PCT/US02/27144

(ABI). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (APB) or PRISM 373 or 377 sequencing systems (ABI) in conjunction with standard protocols and base calling software. Reading frames within the cDNA sequences were identified using standard methods (Ausubel, *supra*, unit 7.7).

### 5 III Extension of cDNAs

The cDNAs were extended using the cDNA clone and oligonucleotide primers. One primer was synthesized to initiate 5' extension of the known fragment, and the other, to initiate 3' extension of the known fragment. The initial primers were designed using primer analysis software to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at 10 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 extension was performed more than one time, 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 15 regions of genes. Genomic libraries can be used to obtain regulatory elements extending 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  $Mg^{2+}$ , 20  $(NH_4)_2SO_4$ , and  $\beta$ -mercaptoethanol, Taq DNA polymerase (APB), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B (Incyte Genomics): The parameters for the cycles are 1: 94C, three min; 2: 94C, 15 sec; 3: 60C, one min; 4: 68C, two min; 5: 2, 3, and 4 repeated 20 times; 6: 68C, five min; and 7: storage at 4C. In the alternative, the parameters for primer pair T7 and SK+ (Stratagene) were as follows: 1: 94C, three min; 25 2: 94C, 15 sec; 3: 57C, one min; 4: 68C, two min; 5: 2, 3, and 4 repeated 20 times; 6: 68C, five min; and 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 Life Sciences, Acton MA) and allowing the DNA 30 to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki Finland) 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 minigel to determine which reactions were successful in extending the sequence.

The extended clones were desalted, concentrated, transferred to 384-well plates, digested with 35 CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared

WO 03/018768

PCT/US02/27144

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

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

#### IV Homology Searching of cDNA Clones and Their Deduced Proteins

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

As detailed in Karim and Altschul (1993; Proc Natl Acad Sci 90:5873-5877), BLAST matches between a query sequence and a database sequence were evaluated statistically and only reported when they satisfied the threshold of  $10^{-25}$  for nucleotides and  $10^{-14}$  for peptides. Homology was also evaluated by product score calculated as follows: the % nucleotide or amino acid identity [between the query and reference sequences] in BLAST is multiplied by the % maximum possible BLAST score [based on the lengths of query and reference sequences] and then divided by 100. In comparison with hybridization procedures used in the laboratory, the stringency for an exact match was set from a lower limit of about 40 (with 1-2% error due to uncalled bases) to a 100% match of about 70.

The BLAST software suite (NCBI, Bethesda MD), includes various sequence analysis programs including "blastn" that is used to align nucleotide sequences and BLAST2 that is used for direct pairwise

WO 03/018768

PCT/US02/27144

comparison of either nucleotide or amino acid sequences. BLAST programs are commonly used with gap and other parameters set to default settings, e.g.: Matrix: BLOSUM62; Reward for match: 1; Penalty for mismatch: -2; Open Gap: 5 and Extension Gap: 2 penalties; Gap x drop-off: 50; Expect: 10; Word Size: 11; and Filter: on. Identity is measured over the entire length of a sequence. Brenner (supra) analyzed BLAST for its ability to identify structural homologs by sequence identity and found 30% identity is a reliable threshold for sequence alignments of at least 150 residues and 40%, for alignments of at least 70 residues.

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

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

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

WO 03/018768

PCT/US02/27144

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). 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.

#### 10 V Northern Analysis, Transcript Imaging, and Guilt-By-Association

##### Northern analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. The technique is described in EXAMPLE VII below and in Ausubel, *supra*, units 4.1-4.9)

Analogous computer techniques applying BLAST are used to search for identical or related molecules in nucleotide databases such as GenBank or the LIFESEQ database (Incyte Genomics). This analysis is faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous. The basis of the search is the product score which was described above in EXAMPLE IV.

The results of northern analysis are reported as a list of libraries in which the transcript encoding TMDC occurs. Abundance and percent abundance are also reported. Abundance directly reflects the number of times a particular transcript is represented in a cDNA library, and percent abundance is abundance divided by the total number of sequences examined in the cDNA library.

##### 25 Transcript Imaging

A transcript image was performed using the LIFESEQ GOLD database (Incyte Genomics). This process allows assessment of the relative abundance of the expressed polynucleotides in all of the cDNA libraries and was described in USPN 5,840,484, incorporated herein by reference. All sequences and cDNA libraries in the LIFESEQ database are categorized by system, organ/tissue and cell type. The categories include cardiovascular system, connective tissue, digestive system, embryonic structures, endocrine system, exocrine glands, female and male genitalia, germ cells, hemic/immune system, liver, musculoskeletal system, nervous system, pancreas, respiratory system, sense organs, skin, stomatognathic system, unclassified/mixed, and the urinary tract. Criteria for transcript imaging are selected from category, number of cDNAs per library, library description, disease indication, clinical relevance of sample, and the like.

WO 03/018768

PCT/US02/27144

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. For each library, the number of cDNAs were counted and shown over the total number of cDNAs in that library. In some transcript images, all enriched, normalized or subtracted libraries, which have high copy number sequences can be removed prior to processing, and all mixed or pooled tissues, which are considered non-specific in that they contain more than one tissue type or more than one subject's tissue, can be excluded from the analysis. Treated and untreated cell lines and/or fetal tissue data can also be excluded where clinical relevance is emphasized. Conversely, fetal tissue can be emphasized wherever elucidation of inherited disorders or differentiation of particular adult or embryonic stem cells into tissues or organs (such as heart, kidney, nerves or pancreas) would be aided by removing clinical samples from the analysis. Transcript imaging can also be used to support data from other methodologies such as hybridization, guilt-by-association and array technologies.

#### Guilt-By-Association

GBA identifies cDNAs that are expressed in a plurality of cDNA libraries relating to a specific disease process, subcellular compartment, cell type, tissue type, or species. The expression patterns of cDNAs with unknown function are compared with the expression patterns of genes having well documented function to determine whether a specified co-expression probability threshold is met. Through this comparison, a subset of the cDNAs having a highly significant co-expression probability with the known genes are identified.

The cDNAs originate from human cDNA libraries from any cell or cell line, tissue, or organ and may be selected from a variety of sequence types including, but not limited to, expressed sequence tags (ESTs), assembled polynucleotides, full length gene coding regions, promoters, introns, enhancers, 5' untranslated regions, and 3' untranslated regions. To have statistically significant analytical results, the cDNAs need to be expressed in at least five cDNA libraries. The number of cDNA libraries whose sequences are analyzed can range from as few as 500 to greater than 10,000.

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

Second, the significance of co-expression is evaluated using a probability method to measure a due-to-chance probability of the co-expression. The probability method can be the Fisher exact test, the chi-squared test, or the kappa test. These tests and examples of their applications are well known in the art and can be found in standard statistics texts (Agresti (1990) Categorical Data Analysis, John Wiley & Sons, New York NY; Rice (1988) Mathematical Statistics and Data Analysis, Duxbury Press, Pacific Grove CA). A Bonferroni correction (Rice, supra, p. 384) can also be applied in combination with one of

WO 03/018768

PCT/US02/27144

the probability methods for correcting statistical results of one gene versus multiple other genes. In a preferred embodiment, the due-to-chance probability is measured by a Fisher exact test, and the threshold of the due-to-chance probability is set preferably to less than 0.001.

This method of estimating the probability for co-expression of two genes assumes that the libraries are independent and are identically sampled. However, in practical situations, the selected cDNA libraries are not entirely independent because: 1) more than one library may be obtained from a single subject or tissue, and 2) different numbers of cDNAs, typically ranging from 5,000 to 10,000, may be sequenced from each library. In addition, since a Fisher exact co-expression probability is calculated for each gene versus every other gene that occurs in at least five libraries, a Bonferroni correction for multiple statistical tests is used (See Walker et al. (1999); Genome Res 9:1198-203; expressly incorporated herein by reference).

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

#### 20 VII Hybridization and Amplification Technologies and Analyses

##### Tissue Sample Preparation

Normal and cancerous tissue samples are described by donor identification number in the table below. The first column shows the donor ID; the second, donor age/sex; the third column, a description of the disorder, the fourth column, classification of the tumor; and the fifth column, the source.

Donor ID	Age/Sex	Tissue and Description	Stage	Source
3579	55/M	colon; well differentiated adenocarcinoma	Dukes' C; TMN T2N1	HCI
3580	38/M	colon; poorly differentiated, metastatic adenoCA	T3N1MX	HCI
3581	U/M	rectal; tumor	NA	HCI
3582	78/M	colon; moderately differentiated adenocarcinoma	TMN T4N2MX	HCI
3583	58/M	colon; tubulovillous adenoma (hyperplastic polyp)	NA	HCI
3647	83/U	colon; invasive moderately differentiated adenocarcinoma (tubular adenoma)	TMN T3N1MX	HCI
3649	86/U	colon; invasive well-differentiated adenoCA	NA	HCI
3479	68/M	colon; adenocarcinoma	NA	HCI
3839	59/M	colon tumor	U	HCI
4614	67/U	colon; moderately differentiated adenocarcinoma	Dukes' B; TMN T3N0	HCI

\*Abbreviations: CA=carcinoma, U=unknown, NA=not available

In Figure 3, the normalized, first-strand synthesis, cDNA preparations of normal, human heart,

WO 03/018768

PCT/US02/27144

brain (whole), lung, liver, skeletal muscle, kidney, pancreas, spleen, thymus, prostate, ovary, small intestine, peripheral blood leukocyte, and colon tissues were obtained from Clontech. Additional cDNA preparations of human, adult, normal thyroid, pituitary, and adrenal tissues were obtained from Clinomics Bioscience (Pittsfield MA).

5 The colorectal adenocarcinoma cell lines shown in Figure 5 were obtained from ATCC and cultured according to the suppliers specifications. The cell lines were, LS123, LS174T, HCT116, CaCo2, HT29, SW480, Colo205, T84, and SW620.

#### Immobilization of cDNAs on a Substrate

10 The cDNAs are applied to a substrate by one of the following methods. A mixture of cDNAs is fractionated by gel electrophoresis and transferred to a nylon membrane by capillary transfer. Alternatively, the cDNAs are individually ligated to a vector and inserted into bacterial host cells to form a library. The cDNAs are then arranged on a substrate by one of the following methods. In the first method, bacterial cells containing individual clones are robotically picked and arranged on a nylon membrane. The membrane is placed on LB agar containing selective agent (carbenicillin, kanamycin, ampicillin, or

15 chloramphenicol depending on the vector used) and incubated at 37C for 16 hr. The membrane is removed from the agar and consecutively placed colony side up in 10% SDS, denaturing solution (1.5 M NaCl, 0.5 M NaOH), neutralizing solution (1.5 M NaCl, 1 M Tris, pH 8.0), and twice in 2xSSC for 10 min each. The membrane is then UV irradiated in a STRATALINKER UV-crosslinker (Stratagene).

In the second method, cDNAs are amplified from bacterial vectors by thirty cycles of PCR using

20 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. 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.

25 Purified nucleic acids are robotically arranged and immobilized on polymer-coated glass slides using the procedure described in USPN 5,807,522. Polymer-coated slides are prepared by cleaning glass microscope slides (Corning Life Sciences) 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

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

35 and rinsed in distilled water as before.

WO 03/018768

PCT/US02/27144

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 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 membrane-based hybridizations as described below.

Probe Preparation for QPCR

Probes for the QPCR were prepared according to the ABI protocol.

Probe Preparation for Polymer Coated Slide Hybridization

The following method was used for the preparation of probes for the microarray analysis presented in Fig. 3. 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 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 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, 1:1,000, and 1:100 (w/w) to sample mRNA respectively. To examine mRNA differential expression patterns, a second set of control mRNAs are diluted into reverse transcription reaction mixture at ratios of 1:3, 3:1, 1:10, 10:1, 1:25, and 25:1 (w/w). The reaction mixture is mixed and incubated at 37C for two hr. The reaction mixture is then incubated for 20 min at 85C, and probes are purified using two successive CHROMA SPIN+TE 30 columns (Clontech, Palo Alto CA). Purified probe is ethanol precipitated by diluting probe to 90  $\mu$ 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.

In situ Hybridization

*In situ* hybridization was used to determine the expression of transmembrane protein in sectioned tissue. With the digoxigenin protocol, fresh cryosections, 10 microns thick, were removed from the freezer, immediately immersed in 4% paraformaldehyde for 10 min, rinsed in PBS, and acetylated in 0.1

WO 03/018768

PCT/US02/27144

M TEA, pH 8.0, containing 0.25% (v/v) acetic anhydride. After the tissue equilibrated in 5 x SSC, it was prehybridized in hybridization buffer (50% formamide, 5 x SSC, 1 x Denhardt's solution, 10% dextran sulfate, and 1 mg/ml herring sperm DNA).

Digoxigenin-labeled TMDC-specific RNA probes, sense and antisense nucleotides selected from the cDNA of SEQ ID NO:2, were produced as follows: 1) a pINCY plasmid containing a fragment of SEQ ID NO:2 extending from about nucleotide 1068 to about nucleotide 2324 of SEQ ID NO:2 (1519 bp) was linearized with EcoRI (antisense) or NotI (sense probe), 2) *in vitro* transcribed using T7 (antisense) or SP6 (sense) RNA polymerase, and 3) hydrolyzed to an average length of 350 bp. Approximately 500 ng/ml of RNA probe was used in overnight hybridizations at 65C in hybridization buffer. Following hybridization, the sections were rinsed for 30 min in 2 x SSC at room temperature, 1 hr in 2 x SSC at 65C, and 1 hr in 0.1 x SSC at 65C. The sections were equilibrated in PBS, blocked for 30 min in 10% DIG kit blocker (Roche Molecular Biochemicals, Indianapolis IN) in PBS, then incubated overnight at 4C in 1:500 anti-DIG-AP. The following day, the sections were rinsed in PBS, equilibrated in detection buffer (0.1 M Tris, 0.1 M NaCl, 50 mM MgCl<sub>2</sub>, pH 9.5), and then incubated in detection buffer containing 0.175 mg/ml NBT and 0.35 mg/ml BCIP. The reaction was terminated in TE, pH 8. Tissue sections were counterstained with 1 µg/ml DAPI and mounted in VECTASHIELD (Vector Laboratory, Burlingame CA).

#### 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 is washed for 15 min at 25C in 1mM Tris (pH 8.0), 1% Sarkosyl, and four times for 15 min each at 25C in 1mM Tris (pH 8.0). To detect hybridization complexes, XOMAT-AR film (Eastman Kodak, Rochester NY) is exposed to the membrane overnight at -70C, developed, and examined visually.

#### Polymer Coated Slide-based Hybridization

The following method was used in the microarray analysis presented in Table 3. Probe is heated to 65C for five min, centrifuged five min at 9400 rpm in a 5415C microcentrifuge (Eppendorf Scientific, Westbury NY), and then 18 µl is aliquoted onto the array surface and covered with a coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5xSSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hr at 60C. The arrays are washed for 10 min at 45C in 1xSSC, 0.1% SDS, and three times for 10 min each at 45C in 0.1xSSC, and dried.

Hybridization reactions are performed in absolute or differential hybridization formats. In the absolute hybridization format, probe from one sample is hybridized to array elements, and signals are

WO 03/018768

PCT/US02/27144

detected after hybridization complexes form. Signal strength correlates with probe mRNA levels in the sample. In the differential hybridization format, differential expression of a set of genes in two biological samples is analyzed. Probes from the two samples are prepared and labeled with different labeling moieties. A mixture of the two labeled probes is hybridized to the array elements, and signals are examined under conditions in which the emissions from the two different labels are individually detectable. Elements on the array that are hybridized to equal numbers of probes derived from both biological samples give a distinct combined fluorescence (Shalon WO95/35505).

Hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective with a resolution of 20 micrometers. In the differential hybridization format, the two fluorophores are sequentially excited by the laser. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, 15 Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Filters positioned between the array and the photomultiplier tubes are used to separate the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. The sensitivity of the scans is calibrated using the signal intensity generated by the yeast control mRNAs added to the probe mix. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the 20 signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the 25 data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using the emission spectrum for each fluorophore. A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the 30 average intensity of the signal. The software used for signal analysis is the GEMTOOLS program (Incyte Genomics).

#### QPCR Analysis

For QPCR, cDNA was synthesized from 1 ug total RNA in a 25 ul reaction with 100 units M-MLV reverse transcriptase (Ambion, Austin TX), 0.5 mM dNTPs (Epicentre, Madison WI), and 40 ng/ml 35 random hexamers (Fisher Scientific, Chicago IL). Reactions were incubated at 25C for 10 minutes, 42C

WO 03/018768

PCT/US02/27144

for 50 minutes, and 70C for 15 minutes, diluted to 500 ul, and stored at -30C. Alternatively, normal tissues were purchased from Clontech (Palo Alto CA) and Clinomics. PCR primers and probes (5' 6-FAM-labeled, 3' TAMRA) were designed using PRIMER EXPRESS 1.5 software (ABI) and synthesized by Biosearch Technologies (Novato CA) or ABI.

5 QPCR reactions were performed using an PRISM 7700 detection system (ABI) in 25 ul total volume with 5 ul cDNA template, 1x TAQMAN UNIVERSAL PCR master mix (ABI), 100 nM each PCR primer, 200 nM probe, and 1x VIC-labeled beta-2-microglobulin endogenous control (ABI). Reactions were incubated at 50C for 2 minutes, 95C for 10 minutes, followed by 40 cycles of incubation at 95C for 15 seconds and 60C for 1 minute. Emissions were measured once every cycle, and results were analyzed using SEQUENCE DETECTOR 1.7 software (ABI) and fold differences, relative concentration of mRNA as compared to standards, were calculated using the comparative  $C_t$  method (ABI User Bulletin #2). QPCR was used to produce the data for Figures 3, 4, and 5

#### VIII Complementary Molecules

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

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

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

#### IX Production of Specific Antibodies

The amino acid sequence of TMDC is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity. An appropriate oligopeptide is synthesized and conjugated to KLH (Sigma-Aldrich).

35 Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant, and

WO 03/018768

PCT/US02/27144

the resulting antisera is tested for antipeptide activity by standard ELISA methods. The antisera is also tested for specific recognition of TMDC. Antisera that reacted positively with TMDC is affinity purified on a column containing beaded agarose resin to which the synthetic oligopeptide had been conjugated. The column is equilibrated using 12 mL IMMUNOPURE Gentle Binding buffer (Pierce Chemical, Rockford IL). Three mL of rabbit antisera is combined with one mL of binding buffer and added to the top of the column. The column is capped on the top and bottom, and antisera allowed to bind with gentle shaking at room temperature for 30 min. The column is allowed to settle for 30 min, drained by gravity flow, and washed with 16 mL binding buffer (4 x 4 mL additions of buffer). The antibody is eluted in one mL fractions with IMMUNOPURE Gentle Elution buffer (Pierce), and absorbance at 280 nm is determined. Peak fractions are pooled and dialyzed against 50 mM Tris, pH 7.4, 100 mM NaCl, and 10% glycerol. After dialysis, the concentration of the purified antibody is determined using the BCA assay (Pierce), aliquotted, and frozen.

#### X Immunopurification Using Antibodies

Naturally occurring or recombinantly produced 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 purified protein is collected.

#### XI Western Analysis

##### Electrophoresis and Blotting

Samples containing protein are mixed in 2 x loading buffer, heated to 95 C for 3-5 min, and loaded on 4-12% NUPAGE Bis-Tris precast gel (Invitrogen). Unless indicated, equal amounts of total protein are loaded into each well. The gel is electrophoresced in 1 x MES or MOPS running buffer (Invitrogen) at 200 V for approximately 45 min on an Xcell II apparatus (Invitrogen) until the RAINBOW marker (APB) is resolved, and dye front approaches the bottom of the gel. The gel and its supports are removed from the apparatus and soaked in 1 x transfer buffer (Invitrogen) with 10% methanol for a few minutes; and the PVDF membrane is soaked in 100% methanol for a few seconds to activate it. The membrane, the gel, and supports are placed on the TRANSBLOT SD transfer apparatus (Biorad, Hercules CA) and a constant current of 350 mAmps is applied for 90 min.

##### Conjugation with Antibody and Visualization

After the proteins are transferred to the membrane, it is blocked in 5% (w/v) non-fat dry milk in 1 x phosphate buffered saline (PBS) with 0.1% Tween 20 detergent (blocking buffer) on a rotary shaker for at least 1 hr at room temperature or at 4C overnight. After blocking, the buffer is removed, and 10 ml of

WO 03/018768

PCT/US02/27144

primary antibody in blocking buffer is added and incubated on the rotary shaker for 1 hr at room temperature or overnight at 4C. The membrane is washed 3 x for 10 min each with PBS-Tween (PBST), and secondary antibody, conjugated to horseradish peroxidase, added at a 1:3000 dilution in 10 ml blocking buffer. The membrane and solution are shaken for 30 min at room temperature and then washed 5 three times for 10 min each with PBST.

The wash solution is carefully removed, and the membrane moistened with ECL+ chemiluminescent detection system (APB) and incubated for approximately 5 min. The membrane, protein side down, is placed on BIOMAX M film (Eastman Kodak) and developed for approximately 30 seconds.

## 10 XII Antibody Arrays

### Protein:protein interactions

In an alternative to yeast two hybrid system analysis of proteins, an antibody array can be used to study protein-protein interactions and phosphorylation. A variety of protein ligands are immobilized on a membrane using methods well known in the art. The array is incubated in the presence of cell lysate until 15 protein:antibody complexes are formed. Proteins of interest are identified by exposing the membrane to an antibody specific to the protein of interest. In the alternative, a protein of interest is labeled with digoxigenin (DIG) and exposed to the membrane; then the membrane is exposed to anti-DIG antibody which reveals where the protein of interest forms a complex. The identity of the proteins with which the protein of interest interacts is determined by the position of the protein of interest on the membrane.

### 20 Proteomic Profiles

Antibody arrays can also be used for high-throughput screening of recombinant antibodies. Bacteria containing antibody genes are robotically-picked and gridded at high density (up to 18,342 different double-spotted clones) on a filter. Up to 15 antigens at a time are used to screen for clones to identify those that express binding antibody fragments. These antibody arrays can also be used to identify 25 proteins which are differentially expressed in samples (de Wildt, *supra*)

### **XIII Screening Molecules for Specific Binding with the cDNA or Protein**

The cDNA, or fragments thereof, or the protein, or portions thereof, are labeled with <sup>32</sup>P-dCTP, Cy3-dCTP, or Cy5-dCTP (APB), or with BIODIPY or FITC (Molecular Probes), respectively. Libraries of candidate molecules or compounds previously arranged on a substrate are incubated in the presence of 30 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.

### 35 **XIV Two-Hybrid Screen**

WO 03/018768

PCT/US02/27144

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

#### XV Cell Transformation Assays

##### 25 Colony-formation Assay in Soft Agar

The ability of transformed cells to grow in an anchorage-independent manner is measured by the ability of the cells to form colonies in soft agar (0.35%). The assay is conducted in 12-well culture plates where each well is coated with a solid 0.7% Noble agar (Fisher Scientific, Atlanta GA) in cell growth media. A 3.5% agar solution in PBS is prepared, autoclaved, microwaved and kept liquid in a 55 C water bath with shaking. The agar is diluted 1:5 to 0.7% with an appropriate cell growth media, and 0.5 ml of the diluted agar added to each well of the plate. Culture plates are kept at room temperature for about 15 minutes or until the agar solidifies.

Trypsinized cells are diluted to 200 to 4000 cells/ml in growth medium and 0.25 ml of diluted cells is mixed with 2 ml warm 0.35% agar. The diluted cells are added to a well of the culture plate; duplicate wells are prepared for each cell concentration. The plates are allowed to cool for about 30 min

WO 03/018768

PCT/US02/27144

at room temperature and then transferred to an incubator at 37 C. After a 1-2 week incubation period, colonies are counted under an inverted, phase contrast microscope. Colony forming efficiency is determined as the percentage colonies formed/total number of cells plated.

Apoptosis/Survival Assay

5 The ability of transformed cells to evade apoptosis (programmed cell death) and survive may be measured in an assay in which apoptosis or survival of cultured cells is determined by FACS analysis using a double-staining method with Annexin V and propidium iodide (PI). Annexin V serves as a marker for apoptotic cells by binding to phosphatidyl serine, a cell surface marker for apoptosis. Counterstaining with PI allows differentiation between apoptotic cells, which are Annexin V positive and PI negative, and 10 necrotic cells, which are Annexin V and PI positive. Apoptosis is measured between 0-24 hrs of culture, and cell survival is measured between 24-96 hrs of culture.

Alternatively, the direct effect of a secreted protein, such as HUPAP, on apoptosis/cell survival may be measured in cultured human vascular endothelial cells (HMVEC) following treatment of HMVEC cells with HUPAP, or infection of the cells with a recombinant adenovirus containing the cDNA encoding 15 HUPAP. Apoptosis/survival of the HMVEC cells is measured as described above.

Tissue Invasion and Metastasis Assay

Cell migration and tissue invasion by transformed tumor cells is determined using the BICOAT Angiogenesis system (BD Biosciences, Franklin Lakes NJ) as described by the manufacturer. The assay is carried out in a BD FALCON multiwell insert plate containing an 8  $\mu$ m pore size BD FLUOROBLOK 20 polyethylene terephthalate membrane uniformly coated with a reconstituted BD MATRIGEL basement membrane matrix and inserted into a non-treated multiwell receiver plate. The system provides a barrier to passive diffusion of cells through the membrane but allows active migration by invasive tumor cells. After cells in appropriate culture medium are incubated in the upper portion of the chamber for a suitable period of time, any cells appearing on the underside of the membrane are quantitated. Since the membrane 25 blocks the transmission of light from 490 to 700nm, cells traversing the membrane are detected by their fluorescence which is proportionate to cell number.

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 30 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.

35

WO 03/018768

PCT/US02/27144

Table 1

<u>Tissue Category</u>	<u># cDNAs</u>	<u>Libraries</u>	<u>Abundance</u>	<u>% Abundance</u>
Cardiovascular System	272986	1/74	1	0.0004
Connective Tissue	151678	0/54	0	0.0000
Digestive System	521762	19/155	40	0.0077
Embryonic Structures	108468	0/24	0	0.0000
Endocrine System	233683	0/63	0	0.0000
Exocrine Glands	258383	5/67	7	0.0027
Genitalia, Female	456353	5/117	7	0.0015
Genitalia, Male	463016	12/120	13	0.0028
Germ Cells	48181	0/5	0	0.0000
Hemic and Immune System	725942	1/179	1	0.0001
Liver	115620	1/37	1	0.0009
Musculoskeletal System	162801	0/50	0	0.0000
Nervous System	995533	0/231	0	0.0000
Pancreas	111771	2/25	2	0.0018
Respiratory System	412898	7/96	9	0.0022
Sense Organs	25345	0/10	0	0.0000
Skin	72732	0/18	0	0.0000
Stomatognathic System	14712	0/17	0	0.0000
Unclassified/Mixed	159180	4/22	4	0.0025
Urinary Tract	295517	2/68	2	0.0007
Totals	5606561	59/1432	87	0.0000

WO 03/018768

PCT/US02/27144

TABLE 2

Library ID *	sDNAs	Library Description	Abundance	%Abundance
COLCTUT03	3344	colon tumor, cecum, adenoCA, 70F	3	0.0897
COLNTUT03	5063	colon tumor, sigmoid, adenoCA, 62M, m/COLNNOT16	3	0.0593
COLCDIT01	3508	colon, cecum, benign familial polyposis, 16M	2	0.0570
COLRTUE01	3657	colon tumor, rectum, adenoCA, mv/tubular adenoma, 50M, SRP	2	0.0547
SINIDME01	4188	sm intestine, ileum, chronic inflammation, 29F, 5RP, EF	2	0.0478
STOMTUP02	18951	stomach tumor, adenoCA, poorly differentiated, 3', CGAP	8	0.0422
COLNTUP17	7421	colon tumor, adenoCA, 3', CGAP	2	0.0270

51

\* All normalized or pooled libraries, mixed tissue libraries, and cell lines were excluded from this analysis

Table 3

DE	P1 Description	P2 Description
1.5	Human, Colon, Ntml, mw/AdenoCA, Dnt3579	Human, Colon Tumor, AdenoCA, Dnt3579
1.9	Human, Colon, Ntml, mw/AdenoCA, Dnt3839	Human, Colon Tumor, AdenoCA Dnt3839
2.2	Human, Colon, Ntml, mw/AdenoCA, Dnt3839	Human, Colon Tumor, AdenoCA Dnt3839

WO 03/018768

PCT/US02/27144

What is claimed is:

1. An isolated cDNA encoding a protein selected from:
  - a) an amino acid sequence of SEQ ID NO:1;
  - b) an antigenic epitope of SEQ ID NO:1, and
  - 5 c) an amino acid sequence having at least 90% sequence identity to the amino acid sequence of SEQ ID NO:1.
2. An isolated cDNA comprising a polynucleotide selected from:
  - a) a nucleic acid sequence of SEQ ID NO:2 or a complement of SEQ ID NO:2;
  - b) a fragment of SEQ ID NO:2 or a complement thereof; and
  - 10 c) a polynucleotide having at least 90% sequence identity to the nucleic acid sequence of SEQ ID NO:2, or a complement thereof.
3. An isolated cDNA comprising a polynucleotide having a nucleic acid sequence of SEQ ID NO:2 or a complement of the cDNA.
4. A probe comprising the cDNA of claim 2.
- 15 5. A cell transformed with the cDNA of claim 2.
6. A composition comprising the cDNA of claim 2 and a labeling moiety.
7. An array element comprising the cDNA of claim 2.
8. A substrate upon which the cDNA of claim 2 is immobilized.
9. A vector comprising the cDNA of claim 2.
- 20 10. A host cell comprising the vector of claim 9.
11. A method for using a cDNA to produce a protein, the method comprising:
  - a) culturing the host cell of claim 10 under conditions for protein expression; and
  - b) recovering the protein from the host cell culture.
12. A composition comprising the cDNA of claim 3 and a labeling moiety.
- 25 13. A method for using a cDNA to detect expression of a nucleic acid in a sample comprising:
  - a) hybridizing the composition of claim 6 to nucleic acids of the sample under conditions to form at least one hybridization complex; and
  - b) detecting hybridization complex formation, wherein complex formation indicates expression of the nucleic acid in the sample.
- 30 14. The method of claim 13 further comprising amplifying the nucleic acids of the sample prior to hybridization.
15. The method of claim 13 wherein the composition is attached to a substrate.
16. The method of claim 13 wherein the sample is from colon or stomach.
17. The method of claim 13 wherein complex formation is compared to standards and is diagnostic of a  
15 colon or stomach cancer.

WO 03/018768

PCT/US02/27144

18. A method of using a cDNA to screen a plurality of molecules or compounds, the method comprising:
- a) combining the cDNA of claim 2 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.
- 5
19. The method of claim 18 wherein the molecules or compounds are selected from antisense molecules, artificial chromosome constructions, branched nucleic acids, DNA molecules, enhancers, peptide nucleic acids, peptides, proteins, repressors, RNA molecules, and transcription factors.
20. A method for using a cDNA to assess efficacy of a molecule or compound, the method comprising:
- a) treating a sample containing nucleic acids with the molecule or compound;
  - b) hybridizing the nucleic acids of the sample with the cDNA of claim 2 under conditions for hybridization complex formation;
  - c) determining the amount of complex formation; and
  - d) comparing the amount of complex formation in the treated sample with the amount of complex formation in an untreated sample, wherein a difference in complex formation indicates efficacy of the molecule or compound.
- 15
21. A method for using a cDNA to assess toxicity of a molecule or compound, the method comprising:
- a) treating a sample containing nucleic acids with the molecule or compound;
  - b) hybridizing the nucleic acids with the cDNA of claim 2 under conditions for hybridization complex formation;
  - c) determining the amount of complex formation; and
  - d) comparing the amount of complex formation in the treated sample with the amount of complex formation in an untreated sample, wherein a difference in complex formation indicates the toxicity of the molecule or compound.
- 20
22. A purified protein selected from:
- a) an amino acid sequence of SEQ ID NO:1;
  - b) an antigenic epitope of SEQ ID NO:1, and
  - d) an amino acid sequence having at least 90% sequence identity to the amino acid sequence of SEQ ID NO:1.
- 25
23. A purified protein of claim 22 comprising an amino acid sequence of SEQ ID NO:1
24. A composition comprising the protein of claim 22 and a labeling moiety.
25. A composition comprising the protein of claim 22 and a pharmaceutical carrier.
26. A substrate upon which the protein of claim 22 is immobilized.
27. An array element comprising the protein of claim 22.
- 30
28. A method for detecting expression of a protein having the amino acid sequence of SEQ ID NO:1 in a
- 35

WO 03/018768

PCT/US02/27144

- sample, the method comprising:
- a) performing an assay to determine the amount of the protein of claim 23 in a sample; and
  - b) comparing the amount of protein to standards, thereby detecting expression of the protein in the sample.
- 5 29. The method of claim 28 wherein the assay is selected from antibody arrays, enzyme-linked immunosorbent assays, fluorescence-activated cell sorting, two dimensional-polyacrylamide gel electrophoresis and scintillation counting, radioimmunoassays, and western analysis.
30. The method of claim 28 wherein the sample is from colon, or stomach.
31. The method of claim 28 wherein the protein is differentially expressed when compared with the
- 10 standard and is diagnostic of a colon or stomach cancer.
32. A method for using a protein to screen a plurality of molecules and compounds to identify at least one ligand, the method comprising:
- a) combining the protein of claim 23 with a plurality of molecules and compounds under conditions to allow specific binding; and
  - 15 b) detecting specific binding, thereby identifying a ligand that specifically binds the protein.
33. The method of claim 32 wherein the molecules and compounds are selected from agonists, antagonists, bispecific molecules, DNA molecules, small drug molecules, immunoglobulins, inhibitors, mimetics, multispecific molecules, peptides, peptide nucleic acids, pharmaceutical agent, proteins, and RNA molecules.
- 20 34. A method for using a protein to identify an antibody that specifically binds the protein having the amino acid sequence of SEQ ID NO:1 comprising:
- a) contacting a plurality of antibodies with the protein of claim 23 under conditions to allow specific binding, and
  - b) detecting specific binding between an antibody and the protein, thereby identifying an antibody
- 25 that specifically binds the protein.
35. The method of claim 34, wherein the plurality of antibodies are selected from a polyclonal antibody, a monoclonal antibody, a chimeric antibody, a recombinant antibody, a humanized antibody, a single chain antibody, a Fab fragment, an F(ab')<sub>2</sub> fragment, an Fv fragment; and an antibody-peptide fusion protein.
36. A method of using a protein to prepare and purify a polyclonal antibody comprising:
- 30 a) immunizing a animal with a protein of claim 22 under conditions to elicit an antibody response;
  - b) isolating animal antibodies;
  - c) attaching the protein to a substrate;
  - d) contacting the substrate with isolated antibodies under conditions to allow specific binding to the protein; and
  - 35 e) dissociating the antibodies from the protein, thereby obtaining purified polyclonal antibodies.

WO 03/018768

PCT/US02/27144

37. A method of using a protein to prepare a monoclonal antibody comprising:
- a) immunizing a animal with a protein of claim 22 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
- 5 antibody producing hybridoma cells;
- d) culturing the hybridoma cells; and
  - e) isolating from culture monoclonal antibody that specifically binds the protein.
38. A method for using a protein to diagnose a cancer comprising:
- a) performing an assay to quantify the expression of the protein of claim 23 in a sample; and
- 10 b) comparing the expression of the protein to standards, thereby diagnosing cancer.
39. The method of claim 38 wherein the sample is selected from colon and stomach.
40. The method of claim 38 wherein expression is diagnostic of a colon or stomach cancer.
41. A method for testing a molecule or compound for effectiveness as an agonist comprising:
- a) exposing a sample comprising the protein of claim 23 to the molecule or compound; and
- 15 b) detecting agonist activity in the sample.
42. A method for testing a molecule or compound for effectiveness as an antagonist, the method comprising:
- a) exposing a sample comprising the protein of claim 23 to a molecule or compound; and
  - b) detecting antagonist activity in the sample.
- 20 43. An isolated antibody that specifically binds a protein having the amino acid sequence of SEQ ID NO:1.
44. A polyclonal antibody produced by the method of claim 36.
45. A monoclonal antibody produced by the method of claim 37.
46. A method for using an antibody to detect expression of a protein in a sample, the method comprising:
- 25 a) combining the antibody of claim 43 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.
47. The method of claim 46 wherein the sample is from colon, liver, lung, ovary, and prostate.
- 30 48. The method of claim 46 wherein complex formation is compared with standards and is diagnostic of a colon or stomach cancer.
49. A method for using an antibody to immunopurify a protein comprising:
- a) attaching the antibody of claim 43 to a substrate;
  - b) exposing the antibody to a sample containing protein under conditions to allow antibody:protein
- 35 complexes to form;

WO 03/018768

PCT/US02/27144

- c) dissociating the protein from the complex; and  
d) collecting the purified protein.
50. A composition comprising an antibody of claim 43 and a labeling moiety.
51. A kit comprising the composition of claim 50.
52. An array element comprising the antibody of claim 43.
53. A substrate upon which the antibody of claim 43 is immobilized.
54. A composition comprising an antibody of claim 43 and a pharmaceutical agent.
55. The composition of claim 54 wherein the composition is lyophilized.
56. A method for using a composition to assess efficacy of a molecule or compound, the method comprising:
- 10 a) treating a sample containing protein with a molecule or compound;  
b) contacting the protein in the sample with the composition of claim 54 under conditions for complex formation;  
c) determining the amount of complex formation; and  
15 d) comparing the amount of complex formation in the treated sample with the amount of complex formation in an untreated sample, wherein a difference in complex formation indicates efficacy of the molecule or compound.
57. A method for using a composition to assess toxicity of a molecule or compound, the method comprising:
- 20 a) treating a sample containing protein with a molecule or compound;  
b) contacting the protein in the sample with the composition of claim 54 under conditions for complex formation;  
c) determining the amount of complex formation; and  
25 d) comparing the amount of complex formation in the treated sample with the amount of complex formation in an untreated sample, wherein a difference in complex formation indicates toxicity of the molecule or compound.
58. A method for treating colon or stomach cancer comprising administering to a subject in need of therapeutic intervention the antibody of claim 43.
59. A method for treating colon or stomach cancer comprising administering to a subject in need of therapeutic intervention the antibody of claim 45.
60. A method for treating colon or stomach cancer comprising administering to a subject in need of therapeutic intervention the composition of claim 54.
61. A method for delivering a therapeutic agent to a cell comprising:
- 35 a) attaching the therapeutic agent to a bispecific molecule identified by the method of claim 33; and  
b) administering the bispecific molecule to a subject in need of therapeutic intervention, wherein the

WO 03/018768

PCT/US02/27144

- bispecific molecule specifically binds the protein having the amino acid sequence of SEQ ID NO:1 thereby delivering the therapeutic agent to the cell.
62. The method of claim 61, wherein the cell is an epithelial cell of the colon.
63. An agonist that specifically binds the protein of claim 23.
- 5 64. A composition comprising an agonist of claim 63 and a pharmaceutical carrier.
65. An antagonist that specifically binds the protein of claim 23.
66. A composition comprising the antagonist of claim 65 and a pharmaceutical carrier.
67. A pharmaceutical agent that specifically binds the protein of claim 23.
68. A composition comprising the pharmaceutical agent of claim 67 and a pharmaceutical carrier.
- 10 69. A small drug molecule that specifically binds the protein of claim 23.
70. A composition comprising the small drug molecule of claim 69 and a pharmaceutical carrier.
71. An antisense molecule of 18 to 30 nucleotides in length that specifically binds a portion of a polynucleotide having a nucleic acid sequence of SEQ ID NO:2 or the complement thereof wherein the antisense molecule inhibits expression of the protein encoded by the polynucleotide.
- 15 72. The antisense molecule of claim 71 wherein the antisense molecule comprises at least one modified internucleoside linkage.
73. The antisense molecule of claim 72 wherein the modified internucleoside linkage is a phosphorothioate linkage.
74. The antisense molecule of claim 71 wherein the antisense molecule comprises at least one nucleotide  
20 analog.
75. The antisense molecule of claim 72 wherein the nucleotide analog is a 5-methylcytidine.

WO 03/018768

PCT/US02/27144

1/14

5' ATG CTG TCC GAT GAC CAC GAG AAT GAA ATC ATC ATA CAG GTT CAG AAT GTT TCC 18 27 36 45 54  
 M L S D D H V N E I I I Q V E N V S  
 TCT GGG GTC CAA AGC CAC CCA TCC TCA AAT CAG AAT TTT CAA GAA AAG GTG CTG 63 72 81 90 99 108  
 S G V Q S H P S S N Q I F Q E K V L  
 CTA GAC TCA AGC ATC AAC ATG GTT TTG TCA ATA TCT GAC AAT CAT GTG ATA GAC 117 126 135 144 153 162  
 L D S S I N M V L S I S D I D V I D  
 TCT CAG ACA GTC AGC AAA AGG AAT GAC CAA AAG GGT AAC CAG GTG CTG CGG TTT 171 180 189 198 207 216  
 S Q T V S K R N D Q K G N Q V L R F  
 TCA ACA TCT TTG AAT GAG TCG ATG TCT CAG ACC CTT CAT AGC CTA GAA TGC ATG 225 234 243 252 261 270  
 S T S L N E S M S Q T L H S L E C M  
 GGC ATA GAC ACT CCT GGT TCT TCA CAT GAA ACT GTT CAA GGA CAG AAG TTA ATC 279 288 297 306 315 324  
 G I D T P G S S H E T V Q G Q K L I  
 GCA TCC CTT ATA CCC ATC ACA TCC AGA GAC AGA AAT AAA GCC ATC AGG AAC CAG 333 342 351 360 369 378  
 A S L I I P M T S R D R I K A I R N Q

FIGURE 1A

387 CCA AGG ACC ATG GAA GAG AAA AGG AAC CTT AGG AAA ATA GTT GAC AAA GAA AAA 432  
 P R T M E E K R N L R K I V D K E K  
 441 AGC AAA CAG ACC CAT CGT ATC CTT CAG CTC AAT TGC TGT ATT CAG TGT CTG AAC 486  
 S K Q T H R I L Q L N C C I Q C L N  
 495 TTC AIT TCC CGG GCT TAT CGG AGA TCC AAG AAC AGC CTG TCG GAA AAT CTG AAT 540  
 S I S R A Y R R S K N S L S E I L N  
 549 TTC AIT AGC CTG TGG CAG AAG ACG CTG AAG ATC AAT GGA GGC AAG TTT GGA ACC 594  
 S I S L W Q K T L K I I G G K F G T  
 603 AGC GTC CTC TCC TAT TTC AAC TTT CTG AGA TGG CTT TTG AAG TTC AAC AAT TTC 648  
 S V L S Y F N F L R W L L K F N I F  
 657 TCA TTC ATC CTG AAC TTC AGC TTC ATC ATA ATC CCT CAG TTT ACC GTG GCC AAA 702  
 S F I L N F S F I I I P Q F T V A K  
 711 AAG AAC ACC CTC CAG TTC ACT GGG CTG GAG TTT TTC ACT GGG GTG GGT TAT TTT 756  
 K N T L Q F T G L E F F T G V G Y F

FIGURE 1B

WO 03/018768

PCT/US02/27144

3/14

765 774 783 792 801 810  
AGG GAC ACA GTG ATG TAC TAT GGC TTT TAC ACC AAT TCC ACC ATC CAG CAC GGG  
R D T V M Y G F Y T N S T I Q H G  
819 828 837 846 855 864  
AAC AGC GGG GCA TCC TAC AAC ATG CAG CTG GCC TFC ATC TTC ACA ATC GGA GCA  
N S G A S Y N M Q L A Y I F T I G A  
873 882 891 900 909 918  
TGC TTG ACC ACC TGC TTC TTC AGT TTG CTG TTC AGC ATG GCC AAG TAT TTC CCG  
C L T T C F F S L L F S M A K Y F R  
927 936 945 954 963 972  
AAC AAC TTC ATT AAT CCC CAC ATT TAC TCC GGA GGG ATC ACC AAG CTG ATC TTT  
N N F I N P H I Y S G G I T K L I F  
981 990 999 1008 1017 1026  
TGC TGG GAC TTC ACT GTC ACT CAT GAA AAA GCT GTG AAG CTA AAA CAG AAG AAT  
C W D F T V T H E K A V K L K Q K N  
1035 1044 1053 1062 1071 1080  
CTT AGC ACT GAG ATA AGG GAG AAC CTG TCA GAG CTC CGT CAG GAG AAT TCC AAG  
L S T E I R E N L S E L R Q E N S K  
1089 1098 1107 1116 1125 1134  
TTG ACG TTC AAT CAG CTG CTG ACC CGC TTC TCT GCC TAC ATG GTA GCC TGG GTT  
L T F N Q L L T R F S A Y M V A W V

FIGURE 1C

4/14

1143 1152 1161 1170 1179 1188  
 GTC TCT ACA GGA GTG GCC ATA GGC TGC TGT CGA GCC GTT TAT TAC CTG GCT GAG  
 V S T G V A I A C C A A V Y Y L A E  
  
 1197 1206 1215 1224 1233 1242  
 TAC AAC TTA GAG TTC CTG AAG ACA CAC AGT AAC CCT GGG CCG GCG CTG TTA CTG  
 Y N L E F L K T H S N P G A V L L L  
  
 1251 1260 1269 1278 1287 1296  
 CCT TTC GTT GTG TCC TGC APT AAT CTG GCC CCA TGC ATC TAC TCC ATG TTC  
 P F V V S C I N L A V P C I Y S M F  
  
 1305 1314 1323 1332 1341 1350  
 AGG CTT GTG GAG AGG TAC GAG ATG CCA CGG CAC GAA GTC TAC GTT CTC CTG ATC  
 R L V E R Y E M P R H E V Y V L L I  
  
 1359 1368 1377 1386 1395 1404  
 CGA AAC ATC TTT TTG AAA ATA TCA ATC ATT GGC APT CTT TGT TAC TAT TGG CTC  
 R N I F L K I S I I G I L C Y Y W L  
  
 1413 1422 1431 1440 1449 1458  
 AAC ACC GTG GCC CTG TCT GGT GAA GAG TGT TCG GAA ACC CTC ATT GGC CAG GAC  
 N T V A L S G E E C W E T L I G Q D  
  
 1467 1476 1485 1494 1503 1512  
 ATC TAC CGG CTC CTT CTG ATG GAT TTT GTG TTC TCT TTA GTC AAT TCC TTC CTG  
 I Y R L L L M D F V F S L V N S F L

FIGURE ID

1521 1530 1539 1548 1557 1566  
 GGG GAG TTT CTG AGG AGA ATC ATT GGG ATG CAA CTG ATC ACA AGT CTT GGC CTT  
 G E F L R R I I G M Q L I T S L G L  
 1575 1584 1593 1602 1611 1620  
 CAG GAG TTT GAC ATT GCC AGG AAC GTT CTA GAA CTG ATC TAT GCA CAA ACT CTG  
 Q E F D I A R N V L E L I Y A Q T L  
 1629 1638 1647 1656 1665 1674  
 GTG TGG ATT GGC ATC TTC TTC TGC CCC CTG CTG CCC TTT ATC CAA ATG ATT-ATG  
 V W I G I F F C P L L P F I Q M I M  
 1683 1692 1701 1710 1719 1728  
 CTT TTC ATC ATG TTC TAC TCG AAA AAT ATC AGC CTG ATG ATG AAT TTC CAG CCT  
 L F I M F Y S K N I S L M M N F Q P  
 1737 1746 1755 1764 1773 1782  
 CCG AGC AAA GCC TGG CGG GCC TCA CAG ATG ATG ACT TTC TTC ATC TTC TTG CTC  
 P S K A W R A S Q M M T F F I F L L  
 1791 1800 1809 1818 1827 1836  
 TTT TTC CCA TCC TTC ACC GGG GTC TTG TGC ACC CTG GCC ATC ACC ATC TGG AGA  
 F F P S F T G V L C T L A I T I W R  
 1845 1854 1863 1872 1881 1890  
 TTG AAG CCT TCA GCT GAC TGT GGC CCT TTT CGA GGT CTG CCT CTC TTC ATT CAC  
 L K P S A D C G P F R G L P L F I H

FIGURE 1E

WO 03/018768

PCT/US02/27144

6/14

1899 1908 1917 1926 1935 1944  
 TTC ATC TAC AGC TGG ATC GAC ACC CTA AGT ACA CGG CCT GGC TAC CTG TGG GGT  
 S I Y S W I D T L S T R P G Y L W V  
  
 1953 1962 1971 1980 1989 1998  
 GTT TGG ATC TAT CCG AAC CTC ATT GGA AGT GTG CAC TTC TTT TTC ATC CTC ACC  
 V W I Y R N L I G S V H F F I L T  
  
 2007 2016 2025 2034 2043 2052  
 CTC ATT GTG CTA ATC ATC ACC TAT CTT TAC TGG CAG ATC ACA GAG GGA AGG AAG  
 L I V L I I T Y L Y W Q I T E G R K  
  
 2061 2070 2079 2088 2097 2106  
 ATT AUG ATA AGG CTG CTC CAT GAG CAG ATC ATT AAT GAG GGC AAA GAT AAA ATG  
 I M I R L L H E Q I I N E G K D K M  
  
 2115 2124 2133 2142 2151 2160  
 TTC CTG ATA GAA AAA TTG ATC AAG CTG CAG GAT ATG GAG AAG AAA GCA AAC CCC  
 F L I E K L I K L Q D M E K K A N P  
  
 2169 2178 2187 2196 2205 2214  
 AGC TCA CTT GTT CTG GAA AGG AGA GAG GTG GAG CAA CAA GGC TTT TTG CAT TTG  
 S S L V L E R R E V E Q Q G F L H L  
  
 2223 2232 2241 2250 2259 2268  
 GGG GAA CAT GAT GGC AGT CTT CAC TTG CGA TCT AGA AGA TCA GTT CAA GAA GGT  
 G E H D G S L D L R S R R S V Q E G

FIGURE 1F

WO 03/018768

PCT/US02/27144

7/14

2277 2286 2295 2304 2313 2322  
 AAT CCA AGG GCC TGA TGA CTC TTT TGG TAA CCA GAC ACC AAT CAA ATA AGG GGA  
 N P R A  
 2331 2340 2349 2358 2367 2376  
 GGA GAT GAA AAT GGA ATG AAT TCT TCC ATG CCA CCT GAG CCT TTA GGA ACT GCC  
 2385 2394 2403 2412 2421 2430  
 CAG AAG AAA ATC CAA GCC TTT AGC CAG GAG CGG AAA CTG ACT ACC ATG TAA TTA  
 2439 2448 2457 2466 2475 2484  
 TCA AAG TAA AAT TGG GCA TTC CAT GCT AAT TTT AAT ACC TGG AAT GCT GAT TTT  
 2493 2502 2511 2520 2529 2538  
 TCA AGA CAA AAT ACT TGG GGT TTT CCA ATA AAG AAT GTT GTA ATA TTG AAA TGA  
 2547 2556 2565 2574 2583 2592  
 GCC TAC AAA AAC CTA GGA AGA GAT AAC TAG GGA ATA ATG TAT ATT ATC TTC AAG  
 2601 2610 2619 2628 2637 2646  
 AAA TGT GTG CAG GAA TGA TTG GTT CTT AGA AAT CTC TCC TGC CAG ACT TCC CAG  
 2655 2664 2673 2682 2691 2700  
 ACC TGG CAA AGG TTT AGA AAC TGT TGC TAA GAA AAG TGG TCC ATC CTG AAT AAA  
 2709 2718 2727 2736 2745 2754  
 CAT GTA ATA CTC CAG CAG GGA TAT GAA GCC TCT GAA TTG TAG AAC CTG CAT TTA

FIGURE 1G

WO 03/018768

PCT/US02/27144

8/14

2763 TTT GTG ACT TTG AAC TAA AGA CAT CCC 2772 2781 2790 2799 2808  
 TGT CCC AAA GGT GGA ATA CAA CCA  
 2817 GAG GTC TCA TCT CTG AAC TTT CTT GCG TAC TGA TTA CAT GAG TCT TTG GAG TCG 2862  
 2871 GGG ATG GAG GAG GTT CTG CCC CTG TGA GGT GTT ATA CAT GAC CAT CAA AGT CCT 2916  
 2925 ACG TCA AGC TAG CTT TGC ACA GAG GCA GTA CCG TAG CCA ATG AGA TTT ATC CGA 2970  
 2979 GAC GCG ATT ATT GCT AAT TGG AAA TTT TCC CAA TAC CCC ACC GTG ATG ACT TGA 3024  
 3033 AAT ATA ATC AGC GCT GGC AAT TTT TGA CAG TCT CTA CGG AGA CTG AAT AAG AAA 3078  
 3087 AAA GAA AAG AAA AGA AAT TAG CTG GGT GCG ATG GCT TAT GCC TGT AAT CCC GGC 3132  
 3141 ACT TTG GGA GGC TGA GGC AAG CGG ATC ACT TAA TGT CAG GAG TTC AAG ACC AGC 3186  
 3195 CTG GCC AAC ATG GTG AAA CCC CGT CTC TAC TAA GGA TAA AAA AAC TGG CTG GGC 3240  
 3249

FIGURE 1H

GTT GTG GTA CAT GCC T 3'

9/14

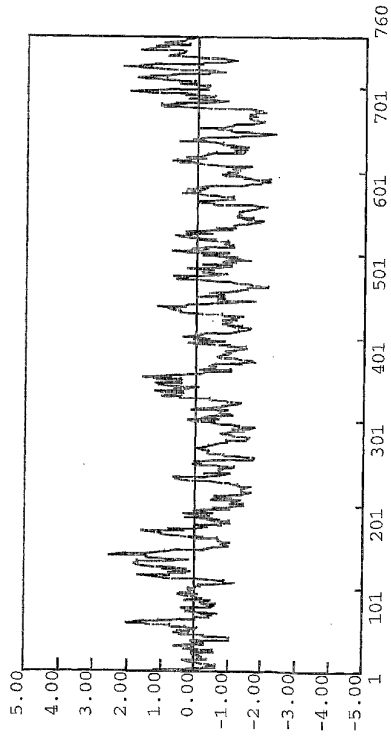


FIGURE 2

10/14

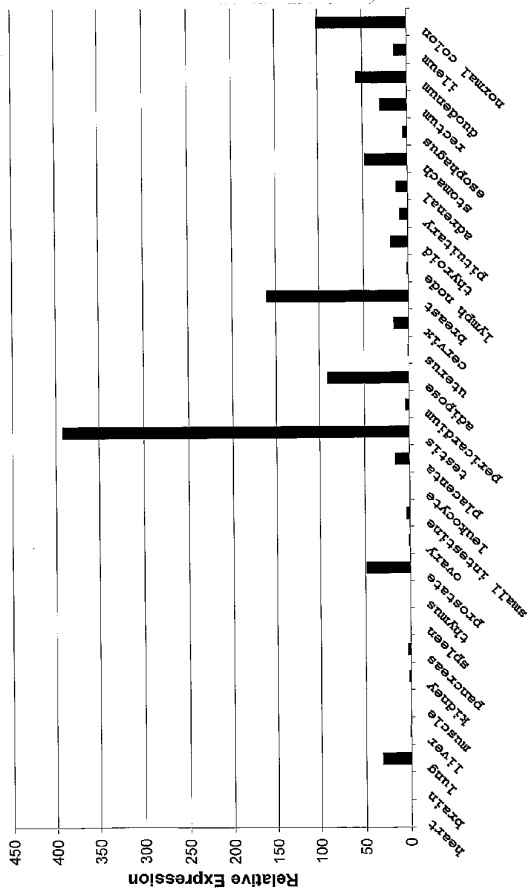


FIGURE 3

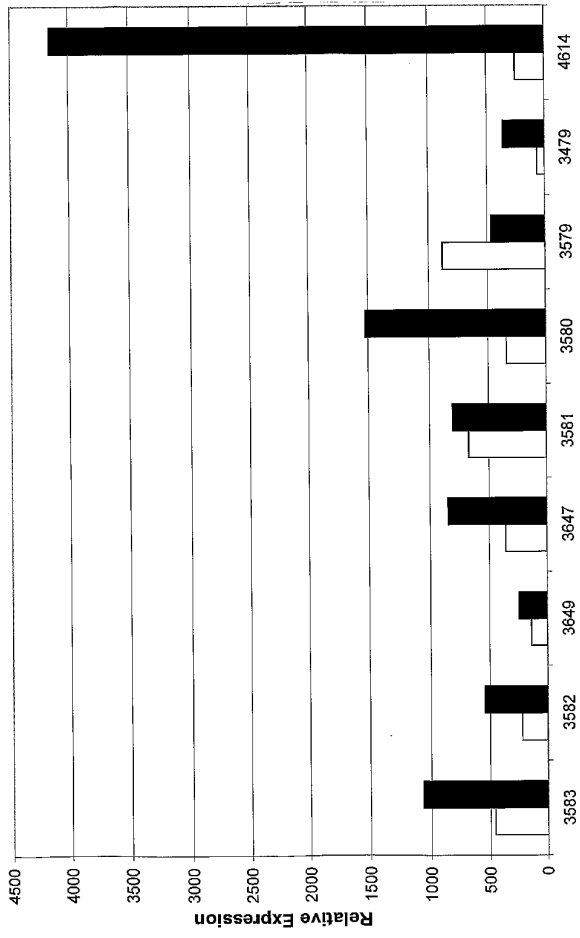


FIGURE 4

12/14

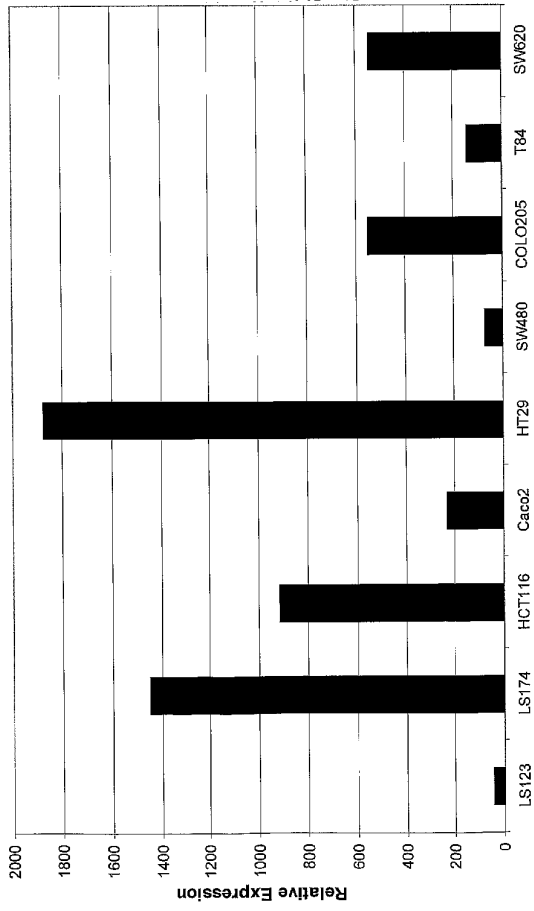


FIGURE 5

WO 03/018768

PCT/US02/27144

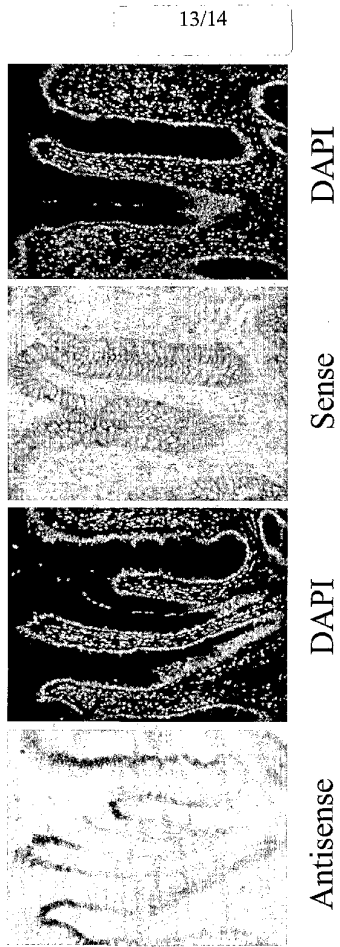


FIGURE 6

14/14

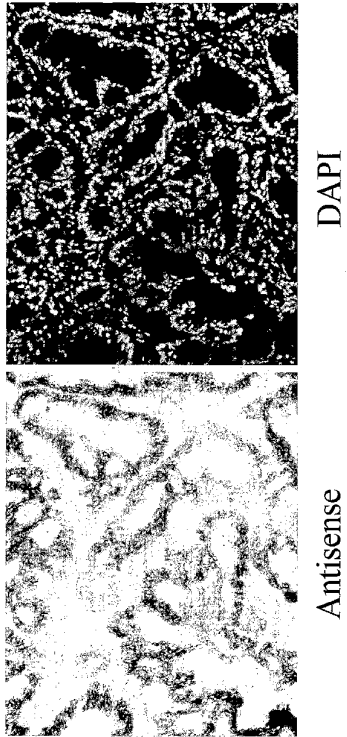


FIGURE 7

WO 03/018768

PCT/US02/27144

<110> INCYTE GENOMICS, INC.  
TRIBOULEY, Catherine M.  
LASEK, Amy K. W.  
YUE, Henry  
BAUGHN, Mariah R.

<120> TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER

<130> PV-0001 PCT

<140> To Be Assigned  
<141> Herewith

<150> US 60/314,914  
<151> 2001-08-24

<160> 16

<170> PERL Program

<210> 1  
<211> 760  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 600000001CD1

```

<400> 1
Met Leu Ser Asp Asp His Val Asn Glu Ile Ile Ile Gln Val Glu
 1                    5                    10
Asn Val Ser Ser Gly Val Gln Ser His Pro Ser Ser Asn Gln Ile
 20                    25
Phe Gln Glu Lys Val Leu Leu Asp Ser Ser Ile Asn Met Val Leu
 35                    40
Ser Ile Ser Asp Ile Asp Val Ile Asp Ser Gln Thr Val Ser Lys
 50                    55
Arg Asn Asp Gln Lys Gly Asn Gln Val Leu Arg Phe Ser Thr Ser
 65                    70
Leu Asn Glu Ser Met Ser Gln Thr Leu His Ser Leu Glu Cys Met
 80                    85
Gly Ile Asp Thr Pro Gly Ser Ser His Glu Thr Val Gln Gly Gln
 95                    100
Lys Leu Ile Ala Ser Leu Ile Pro Met Thr Ser Arg Asp Arg Ile
110                    115
Lys Ala Ile Arg Asn Gln Pro Arg Thr Met Glu Glu Lys Arg Asn
125                    130
Leu Arg Lys Ile Val Asp Lys Glu Lys Ser Lys Gln Thr His Arg
140                    145
Ile Leu Gln Leu Asn Cys Cys Ile Gln Cys Leu Asn Ser Ile Ser
155                    160
Arg Ala Tyr Arg Arg Ser Lys Asn Ser Leu Ser Glu Ile Leu Asn
170                    175
Ser Ile Ser Leu Trp Gln Lys Thr Leu Lys Ile Ile Gly Gly Lys
185                    190
Phe Gly Thr Ser Val Leu Ser Tyr Phe Asn Phe Leu Arg Trp Leu
200                    205
Leu Lys Phe Asn Ile Phe Ser Phe Ile Leu Asn Phe Ser Phe Ile
215                    220
Ile Ile Pro Gln Phe Thr Val Ala Lys Lys Asn Thr Leu Gln Phe
230                    235
Thr Gly Leu Glu Phe Phe Thr Gly Val Gly Tyr Phe Arg Asp Thr

```

WO 03/018768

PCT/US02/27144

	245		250		255
Val Met Tyr Tyr	Gly Phe Tyr Thr Asn Ser Thr Ile Gln His Gly				
260			265		270
Asn Ser Gly Ala Ser	Tyr Asn Met Gln Leu Ala Tyr Ile Phe Thr				
275			280		285
Ile Gly Ala Cys Leu	Thr Thr Cys Phe Phe Ser Leu Leu Phe Ser				
290			295		300
Met Ala Lys Tyr Phe	Arg Asn Asn Phe Ile Asn Pro His Ile Tyr				
305			310		315
Ser Gly Gly Ile Thr	Lys Leu Ile Phe Cys Trp Asp Phe Thr Val				
320			325		330
Thr His Glu Lys Ala	Val Lys Leu Lys Gln Lys Asn Leu Ser Thr				
335			340		345
Glu Ile Arg Glu Asn	Leu Ser Glu Leu Arg Gln Glu Asn Ser Lys				
350			355		360
Leu Thr Phe Asn Gln	Leu Leu Thr Arg Phe Ser Ala Tyr Met Val				
365			370		375
Ala Trp Val Val Ser	Thr Gly Val Ala Ile Ala Cys Cys Ala Ala				
380			385		390
Val Tyr Tyr Leu Ala	Glu Tyr Asn Leu Glu Phe Leu Lys Thr His				
395			400		405
Ser Asn Pro Gly Ala	Val Leu Leu Leu Pro Phe Val Val Ser Cys				
410			415		420
Ile Asn Leu Ala Val	Pro Cys Ile Tyr Ser Met Phe Arg Leu Val				
425			430		435
Glu Arg Tyr Glu Met	Pro Arg His Glu Val Tyr Val Leu Leu Ile				
440			445		450
Arg Asn Ile Phe Leu	Lys Ile Ser Ile Ile Gly Ile Leu Cys Tyr				
455			460		465
Tyr Trp Leu Asn Thr	Val Ala Leu Ser Gly Glu Glu Cys Trp Glu				
470			475		480
Thr Leu Ile Gly Gln	Asp Ile Tyr Arg Leu Leu Met Asp Phe				
485			490		495
Val Phe Ser Leu Val	Asn Ser Phe Leu Gly Glu Phe Leu Arg Arg				
500			505		510
Ile Ile Gly Met Gln	Leu Ile Thr Ser Leu Gly Leu Gln Glu Phe				
515			520		525
Asp Ile Ala Arg Asn	Val Leu Glu Leu Ile Tyr Ala Gln Thr Leu				
530			535		540
Val Trp Ile Gly Ile	Phe Phe Cys Pro Leu Leu Pro Phe Ile Gln				
545			550		555
Met Ile Met Leu Phe	Ile Met Phe Tyr Ser Lys Asn Ile Ser Leu				
560			565		570
Met Met Asn Phe Gln	Pro Pro Ser Lys Ala Trp Arg Ala Ser Gln				
575			580		585
Met Met Thr Phe Phe	Ile Phe Leu Leu Phe Phe Pro Ser Phe Thr				
590			595		600
Gly Val Leu Cys Thr	Leu Ala Ile Thr Ile Trp Arg Leu Lys Pro				
605			610		615
Ser Ala Asp Cys Gly	Pro Phe Arg Gly Leu Pro Leu Phe Ile His				
620			625		630
Ser Ile Tyr Ser Trp	Ile Asp Thr Leu Ser Thr Arg Pro Gly Tyr				
635			640		645
Leu Trp Val Val Trp	Ile Tyr Arg Asn Leu Ile Gly Ser Val His				
650			655		660
Phe Phe Phe Ile Leu	Thr Leu Ile Val Leu Ile Ile Thr Tyr Leu				
665			670		675
Tyr Trp Gln Ile Thr	Glu Gly Arg Lys Ile Met Ile Arg Leu Leu				
680			685		690
His Glu Gln Ile Ile	Asn Glu Gly Lys Asp Lys Met Phe Leu Ile				
695			700		705
Glu Lys Leu Ile Lys	Leu Gln Asp Met Glu Lys Lys Ala Asn Pro				
710			715		720

WO 03/018768

PCT/US02/27144

Ser	Ser	Leu	Val	Leu	Glu	Arg	Arg	Glu	Val	Glu	Gln	Gln	Gly	Phe
				725					730					735
Leu	His	Leu	Gly	Glu	His	Asp	Gly	Ser	Leu	Asp	Leu	Arg	Ser	Arg
				740					745					750
Arg	Ser	Val	Gln	Glu	Gly	Asn	Pro	Arg	Ala					
				755					760					

<210> 2  
 <211> 3256  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 60000001CB1

<400> 2  
 atgctgtccg atgaccacgt gaatgaaatc atcatcacagg ttgagaatgt ttccctctggg 60  
 gtccaaagcc acccaccctc aaatcagatt ttccaagaaa aggtgctgct agactcaagc 120  
 atcaaatcgg ttttgtcaat atctgacatt gatgtgatag actctcagac agtcagcaaa 180  
 aggaatgacc aaaagggtaa ccaggtgctg cggttttcaa catctttgaa tgagtctgat 240  
 tctcagaacc tccatagcct agaatgcatg ggcatagaca ctctcgttc ttccatgaa 300  
 actgttcaag gacagaagtt aatcgcaacc ctataccaca tgacatccag agacagaatt 360  
 aaagccatca gaaacagacc aaggaccatg gaagagaaaa ggaacctag gaaalagtt 420  
 gacaangaaa aaagcaaaaa gaccatcgt atcctcagc tcaattcctg tcttcagctg 480  
 ctgaactcca ttcccgggc ttatcggaga tccaagaaca gctctgga aattctgaat 540  
 tccatcagcc tctgtccaga gacgtgaag atcaltggag gcaagtttg aaccagctc 600  
 ctctccatatt tcaactttct gagatggctt ttgaagtcca acattttctc attcaatcctg 660  
 aacttcaagct tccatcataat ccctcagttt acogtggcca aaaagaacac cctccagctc 720  
 actggctggg agtttttacc tggggtgggt tabtttaggg acacagtgat gtaactatggc 780  
 ttttaacaca attccaccat acagcagggg aacagcgggg catctacaa catgcaactg 840  
 gcctacatct tcaacaatcg agcatcctg accacctgct tctctcagtt gctgttcagc 900  
 atggccaagt atttccggaa caactcatt aatccccaca tttactcgg aggatcacc 960  
 aagctgatct tttgtggga ctccactgtc actcatgaaa aagctgtgaa gctaaaacag 1020  
 aagaatctta gcactgagat aagggagaac ctgtcagagc tccgtcagga gaattccaag 1080  
 ttgacttcca atcagctgct gaccctctc tctgctcaca tggtagcctg ggttctctct 1140  
 acagagtggt ccatagcctg ctctcagacc gtttattacc tggctgagta caacttagag 1200  
 ttcctgagaa cacacagtaa cctctggggc gtgctgttac tgcctttcgt tctgtctcgt 1260  
 ataatctdgg ccgtgccatg catctactcc atgttcaggc ttgtggagag gtaccagatg 1320  
 ccacggcacg aagtctaact tctcctgact ogaaacatct tttgaaaat atcaatcatt 1380  
 ggcattcttt gtlactattg gctcaaaccc gtggccctgt ctggggaaga gtgttgggaa 1440  
 accctcattg gccaggacat ctaccggctc ctctctgatg attttgggtt ctcttttagc 1500  
 aattcctccc tggggagatt tctgaggaga atcattggga tgcaactgat cacaagtctt 1560  
 ggctctcagg agtttgacat tgccaggaac gttctagaac gtgatctatg acaaaactctg 1620  
 gtgtggatgt gcaatctctt ctgcccctgt ctgcccctta tccaaatgat tatgcttttc 1680  
 atcatgttct actcaaaaaa taccagctgt atgatgaatt tccagctccc gagcaagcc 1740  
 tggcgggccc cacagatgat gactttcttc atctcttctg totttttccc atccttcacc 1800  
 ggggtcttctg gcaccctggc catcaccatc tggagattga agccttcagc tgactgtggc 1860  
 ccttttccag gctcgcctct ctccattcac tccatctaca gctggatcga caccctaagl 1920  
 aacaggctct gctaacctgt ggttgtttgg atctatcgga acctcaatgg aagtgtcac 1980  
 ttcttttcca tctccacctc cattgtgcta atcatcact atcttactg gcagatcaca 2040  
 gagggaagga agattatgat aaggctgctc catgagcaga tcaatgatga gggcaaatg 2100  
 aaaaatctcc tgatagaaaa attgatcaag ctgacagata tggagaagaa agcaaacccc 2160  
 agctcaacttc ttctggaaaag gagagagtg gagcaacaag gctttttgca ttgggggaa 2220  
 catgatggca gtcttgactt cggatctaga agatcagttc aagaaggtaa tccaagggcc 2280  
 tgatgactct tttgtaacc agacaccaat caaatagggg gagagatga aatggaatg 2340  
 atttcttcca tgcacacctg gcctttagga actgcccaga agaaaatcca aggttttagc 2400  
 caggagcggg aactgactac catgtaatta tcaaatgaaa atggggcatt ccatgctatt 2460  
 ttaaatcctc gnatgtctga tttttcaaga caaaatactt ggggttttcc aataagatt 2520  
 gttgtaatat tgaatgagc ctacaaaaac ctaggaaag ataaactagg aataatglat 2580  
 atattttcca agaatgtgt gcaggaatga ttggttctta gaaactctc ctgcaagact 2640  
 tccagacctt ggaagagtt tagaactgt tgcataaana agtggctcat cctgaataaa 2700  
 catgtaatac tccagcaggg atatgaagcc tctgaattgt agaacctgca tttatttctg 2760

WO 03/018768

PCT/US02/27144

```

accttgaact aaagacatcc cccatgtccc aaaggtggaa tacaaccaga ggtctcatct 2820
ctggaacttc ttgggtactg attacatgag tctttggagt cggggatgga ggaggtctctg 2880
ccctgtggag gtgttatata tgaccatcaa agtctctact caagctagct ttgcacagtg 2940
gcaataccgt agccaatgag atttatccga gacgagatta ttgctaattg gaaatlttcc 3000
caatacccca cogtgalgac ttgaaatata atcagcgctg gcaatttttg acagtctctca 3060
cggagactga ataagaaaaa agaaaagaaa agaaattagc tgggtgcat ggcttatgcc 3120
tghtaatccc gcaacttggg aggtgagggc aagcggatca cttaatgtca ggagttcaag 3180
accagcctgg ccaacatggt gaaaccccgct ctctactaag gataaaaaaa ctggctgggc 3240
gtgggtggtac atgcoct 3256

```

```

<210> 3
<211> 272
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 1929823H1

```

```

<400> 3
tgaggagaat cattgggatg caactgatca caagctttgg ccttcaggag ttgacattg 60
ccaggaacgt tctagaactg atctatgcac aaactctggt gtggattggc atcttctctc 120
gccccctgct gccctttatc caaatgatga tgcctttcat catgttctac tccaaaaata 180
tcagcctgat gatgaaatcc cagcctccga gcaaaacctg gggggcctca cagatgatga 240
ctttctctat ctctttgctc tttttcccat cc 272

```

```

<210> 4
<211> 413
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 1929823T6

```

```

<220>
<221> unsure
<222> 94, 160, 210, 227, 387
<223> a, t, c, g, or other

```

```

<400> 4
cgctgattat atttcaagtc atcacggtgg ggtattggga aaatttccaa tttagcaata 60
tcgctctctg gataaatctc attggctacg gtantgccac tggcaaaagc tagcttgagc 120
taggactttg atggtcatgt ataacacctc acagggccan aacctctctc atccccgact 180
ccaagaactc atgtaatcag tacgcaagan agttcagaga tgagacntct ggttgbattc 240
caactttggg acatggggga tgtctttagt tcaaagtcac aaataatgc aggttctaca 300
attcagaggg ttcatatccc tgcctggagta ttacatgttt attcaggatg gaccactttt 360
cttagcaaca gtttctaaac ctttgenagg tctggggaag tctgggcagg gag 413

```

```

<210> 5
<211> 497
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 1341151F6

```

```

<220>
<221> unsure
<222> 193, 435, 451
<223> a, t, c, g, or other

```

WO 03/018768

PCT/US02/27144

```

<400> 5
cagatgatga ctttttccat ctttttgcct tttttcccat ccttcaccgg ggtccttgtc 60
acctgtggca ccacoactct gagattgaag ccttcagctg actgtggccc ttttggaggt 120
ctgcctctct ccattcactc catctacagc tggatcgaca cctlaagtac accgcttggc 180
tacctgtggg ttntttggat ctatcggaac ctcaattggaa gtgtgcattc tttttcatcc 240
tcaccctcat tgtgtaaatc atcaccatc tttactggca gatcacagag ggaaggaaga 300
ttatgataag gctgctccat gagcagatca ttaatgaggg caaagataaa atgtcctgat 360
agaaaaattg atcaagctgc aggatatgga gaagaaagca aaccccagct tcacttgttc 420
tgggaaagga gagangtga gcaacaaggc nttttgcatt tgggggaaca tgatgggag 480
tcttgacttg cgattct 497

```

```

<210> 6
<211> 532
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7703595H1

```

```

<400> 6
ggggtgggtt attttaggga cacagtgatg tactatggct tttacaccaa tbcaccatc 60
cagcacggga ccagcggggc atectacaac atgcagctgg cctacatctt cacaatcggc 120
gcctgcttga ccacctgctt cttcagtttg ctgttcagca tggccaagta tttcgggac 180
aacttcattc atcccacat ttactccgga gggatcaca agctgatctt ttgctgggac 240
ttcactgtca ctcatgaaa agctgtgaa gtaaaacaga agaattctag cactgagata 300
agggagaacc tgtcagagct ccctcaggag aattccaagt tgaacttcaa tcagctctg 360
accgctctc ctgcctacat ggtagcctgg gttgtctcta caggagctgg catagctctg 420
tgtcagccg tttattacct ggtgagttac aacttagagt tccctgaagc acacagtaac 480
cctggggcgg tgcgttaact gcctttcgtt gtgtcctgca ttaacttggc cg 532

```

```

<210> 7
<211> 638
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 8146316H1

```

```

<400> 7
ccgggggat caccaagctg atctttgctg ggacttcaat gtcactcatg aaaaagctgt 60
gaagctaaaa cagaagaatc ttagcactga gataaggag aacctgtcag agctccytca 120
ggagaattcc aagttgactt tcaatcagct gctgaccgc ttctctcctt acatggtagc 180
ctgggttctc tctacaggag tggccatagc ctgctgtgca gccgtttatt acctggctga 240
gtacaactta gacttctga agacacacag taacctggg ggggtgctgt tactgccttt 300
cgttgtgtcc tgcattaatc tggccgtgcc atgcactcac tccatgttca ggttgggga 360
gaggtacgag atgccaaggg ccgaagtcta cgttctctct atccgaaaca tcttttggaa 420
aatataatc atggccattc ttgttacta ttggtcaac accgtggccc tgbctgggta 480
agaggttgg gaacacctca ttggccagga catctaccgg ctctcttga tggatttgg 540
ttctcttag tcaattcctt cctgggggag tttctgagga gaatcattg atgcaactga 600
tcacnagctc tggcctcag gatttgaca ttgcnagg 638

```

```

<210> 8
<211> 71
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 3274531H1

```

```

<400> 8

```

WO 03/018768

PCT/US02/27144

caataccocca ccgtgatgac ttgaaatata atcagcgctg gcaatttttg acagtctcta 60  
 cggagactga a 71

<210> 9  
 <211> 540  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: SCCA02331V1

<220>  
 <221> unsure  
 <222> 6-7  
 <223> a, t, c, g, or other

<400> 9  
 gatctnltca tggagcagcc ttatcataat ctccctccc tctgtgatct gccagtaaaag 60  
 ataggtgatg attagcacia tgagggtgag gatgaaaaag aagtgcacac ttccaatgag 120  
 gtccgatag atccaaacaa cccacaggta gccaggccgt gtacttaggg tgcgatcca 180  
 gctgtagatg gactgaatga agagagggcag acctegaaaa gggccacagt cagctgaaag 240  
 ctccaatttc cagatgggtga tggccagggg gcaccagacc ccggtgaaag atgggaaaaa 300  
 gaggcaagaa atgaagaag tcatcatctg cagggcccg caggetttgc tggagggctg 360  
 gaaattcacc atcagctgga tatttttggg gtgaaacatg atgaaaagca taatcaattg 420  
 gataaagggc agcagggggc agaagaagat gccaatccc accagagttt gtccatgat 480  
 cagttctaga acgttctctg caatgtcaac tccctgaagc caagacttgt gatcagttgc 540

<210> 10  
 <211> 567  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: SCCA04417V1

<220>  
 <221> unsure  
 <222> 248, 339, 523, 539, 551  
 <223> a, t, c, g, or other

<400> 10  
 gaatgatttc ttccatgcca cctgtgcctt taggaactgc ccagaagaaa atccaaggct 60  
 ttgcccagga gcggaactg actaccatgt aattatcaaa gtaaaatgg gcattccatg 120  
 ctatttttaa taccctggatt gctgattttt caagacaaaa taactgggggt ttccaataa 180  
 agattgttgt aatattgaaa tgagcctaca aaaacctagg aagagataac tagggaataa 240  
 tgtattnat ctccagagag tctgtgcagg aatgattggt tctlagaat ctctctgcgc 300  
 agaactccca gaactgcaa aggttagaa actggtgema agaaaagtgg tccatccatg 360  
 ataaacatgt gatactccag cagggatag aagcctctga atbtgagaac ctgcatttat 420  
 tttgtgactt tgaacttaaa gacatcccc catgtccaa aggtggaata caaccagagg 480  
 totcatctct gaactttctt cgtctctgat tacatgagtt ttngaggtgg gggatggang 540  
 aggtcttccc ntggtagggg ttaacat 567

<210> 11  
 <211> 2421  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: g2951946\_010

WO 03/018768

PCT/US02/27144

```

<400> 11
atgctgtccg atgaccacgt gaatgaaatc atcatabcagg ttgagaatgt ttccctctggg 60
gtccaagacc acccatcctc aaatcagatt ttccaagaaa aggtgctgct agactcaagc 120
atcaacatgg ttttgtcaat atctgcacat gatgtgatag actctcaagc agtcagcaaa 180
aggaatgaacc aaaaagggtaa ccagggtctg cggttttcaa catctttgaa tgagtcgatg 240
tctcagaccc ttcatagctc agaatgcatg ggcatagaca ctccctggttc ttcacatgaa 300
actgttcaag gacagaagtt aatcgcaccc ctlatacca tgacatccag agacagaatt 360
aaagccatca ggaaccagcc aaggaccatg gaagagaaaa ggaaccttag gaaaatagtt 420
gaccaagaaa aaagcaaaaca gaccatcgt atccttcaag tcaattgctg tattcagttg 480
ctgaactoca tttcccgggc ttatcggaga tccaagaaaca gctctgctgga aattctgaat 540
tccatcaagc ttggccagaa gaogctgaag atcaattggag gcaagtttgg aaccagctc 600
ctctctatt tcaactttcc gagatggctt ttgaagtcca acacttttcc attcaactg 660
aaotcaagct tcaatacaat cccctcagtt accgtggcca aaagaaacc cctccagttc 720
actggctcgg agttttcaac tgggggtggg ttttttaggg acacagtgat gtaactatgg 780
ttttacacca attccacaact ccaggcaggg aaagcgggg catcctacaa catgcaagctg 840
gctacatctc tcaaatcgg agcatgcttg acccctgct tcttaagttt gctgttcaag 900
atggccaagt atttccggaa caacttcatt aatcccaca tttactccgg agggatcaac 960
aagctgatct tttgctggga ettcaactgc actcatgaaa aagctgtgaa gctaaaacag 1020
aagaatctta gcactgagat aaggagaaac ctgctcagag cctgtcagga gaattcaag 1080
ttgagcttca atcagctgct gaccgccttc tctgcctaca tggtagcctg ggtgtctct 1140
acaggatggg ccatagcctg ctgtgcagcc gtttattacc tggctgagta caacttagag 1200
gtacccaaca ccagggtoca gggcagagag aaccagttcc tgaagacaca cagtaacctc 1260
ggggcggctg tgttaactgc tttcgttctg tctgcatta atctggcctg gccatgcatc 1320
tactcaatgt tcaaggctgt ggagaggtac gagatgccc ggaacgaagt ctacgttctc 1380
ctgactcgaa acatcttttt gaaaatata atcaattggca tctttgtta ctattggctc 1440
aaacacctgg cccctgctgg tgaagaggtg ttggaaacc ccaattggca ggaactatc 1500
cgtctcttcc tgaatgatt ttgtttctct ttactcaatt ccttccggg ggagttctg 1560
aggaaatca ctggatgca actgatcaca agtcttggcc ttcaagagtt tgacattgcc 1620
aggaactgtc tagaactgat ctatgcacaa actctgggtg ggattggcat cttcttctg 1680
ccccctgct cctttatcca aatgattatg cttttcatca ttttactc caaaaatgtg 1740
agtcaagtcg acattgccat caatcagatt ttttcaagtc cctgtacct ggtggcctt 1800
aaagctgggg aagggggctc tgaagaatc agcctgatga tgaatttcca gctccagagc 1860
aaagcctggc gggcctcaca gatgatgaat ttcttcaatc tcttgccttt tttccatcc 1920
tcaaccgggg tcttctgctc cctggccatc accatctgga gattgaagcc ttcagctgac 1980
tgggcccctt ttcagagttc gctctcttcc attcaactcca tctacagctg gatgcaacc 2040
ctaagtaaac ggcctggcta cctgtgggtt gtttggatct atcggaacct cactggaagt 2100
gtgacttct ttttcaatcc caccctcatt gtgctaatca tcaactatct ttaactggag 2160
atcacagagg gaaggaagat tatgalaagg ctgtccatg agcagatcat taatggggc 2220
aaagataaaa tgttctgat agaaaaatg atcaagctgc agpatatgga gaagaaagca 2280
aacccagctc cacttgttct ggaaggaga gaggtgagc aacaaggctt ttgcatltg 2340
gggaaacatg atggcagctc tggaaactgc cagaagaaaa tccaaggctt tagccaggag 2400
cggaaactga ctaccatgta a

```

```

<210> 12
<211> 198
<212> DNA
<213> Rattus norvegicus

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 701294553H1

```

```

<400> 12
gcatctgct gtgctcagtg tctcagctcc ctttccctgg cttaccgagg aaccaagagc 60
agccttcaag agctcctcaa ttacatcagc ctgtggcaga agagattcaa ggtcatcgga 120
ggcaagtttg gaaccagctc cctgtcctat ttcagcttcc tgaggtggct ttgaaagttc 180
aacatcttct cattcgtc

```

```

<210> 13
<211> 306
<212> DNA
<213> Rattus norvegicus

```

WO 03/018768

PCT/US02/27144

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 701600294H1

&lt;400&gt; 13

```

ctggaaacaa gttggatttt tttttocaat tagcaacaat cgcacctgg ataacctca 60
ctggctatga lactgocact gtgcaaaagt gttttttttt ttaaccaaag tgactcttac 120
ctactagctc cagaaggggt ggctctggag aggtgcagcc caggaagggt gcctctgtct 180
tggctggaga ttlgacagat tgaacacagc ctctctgatg caaatcagac catggagtc 240
cacactttaa ttcccccaat ttgtcttttt attttacaag gtggaaagcct ccggtgtctc 300
ctctgc 306

```

&lt;210&gt; 14

&lt;211&gt; 156

&lt;212&gt; DNA

&lt;213&gt; Mus musculus

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2016808H1

&lt;400&gt; 14

```

cgagcggccg ccggggcagg tcaaaaattg ccaatgccga ctatattgca agtcgtcaag 60
gcccgggtatt gggaaaagt ttcaatttagc aataatcggg cctcggataa acctcattgg 120
ctacgatact gccaacccgc ctccgcacca cgcct 156

```

&lt;210&gt; 15

&lt;211&gt; 1370

&lt;212&gt; DNA

&lt;213&gt; Mus musculus

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 239780\_Mm.1

&lt;400&gt; 15

```

gttcacatc aggtgaagtt ttgacatag acataatgaa gagatgatca tttggataaa 60
ggcagcaagg gaagaagaga ttccagcca ggtcagatt gtccgtagat cagttctaga 120
acatctcctg caatgtcaac tctctgagct gagactggtg aacttcaccc catgagcctc 180
ctcagaaact ccgcagtaa ggaatcggcc aaggagaaca cgaagtccat gagaggagc 240
cggtagatgt cctggccaat gaggtctccc cagcactctt cgcagacagc ggccacgatg 300
ttgagccaat agtaacaaag aattccaaca atggagatct tcaaaaagat gtttcggacc 360
aagaggcagt agaactcctg tcttggtatc tcatacctct ccaccaggcg gaacatggag 420
tagaagcgag gcaogggcag gttgatgca gacacaacia agggcagcaa caacaccgcc 480
ccagggttcc tgtgagtctt caggaaactca gagttatact cagccaagta gtgacagct 540
acacgcaggg ctgoggtcac tccagtagag acgagccaag cggctacgtg agcagaaaaa 600
cgggtcaagt gctggtttaa tgtagcctg gtatctctcc tggcggagct cagaccaggt 660
tctccctgat ctccggtctc agatctctct gtttagctt tacagctttt tcaagggtga 720
cagtgaaatc ccagcaaaag atgagcttag caatccctct ggagtatatg tggggttga 780
tgaagttgtt ccggaagtag ttgtccatad tgaagagtag actgaagaaa cagacgacca 840
ggcagctccc gatcgtgaag atgtagacca cctgcattgl gtaggatgct ccaaccoatc 900
latgtcggat ttagaattg gtgtgacc catagtaact caccgtctcc caaaaatac 960
ccgcccctgt gcaaaaactc caagccagtg aactggcagg gtagttcttt gcacccacgg 1020
tgaactctgt ggatcgtatg tgaagctgaa gttcactgac gcaatcgcag atacgatgta 1080
tgaactctcc aaacagccacc tcaggacagc tgataatcac ggaacaggac cctcgggttc 1140
aaactctcgc ctccgtatga ccttgaatct ctctcgcacc aagtgatgat aattgaggag 1200
ctctcgaagg tctctgtctc cctcctcctc ggtacaagtc cagtgaaaag gagcccagac 1260
actgcagcac agcaagttgg cctcagggac tgccatgtga ctgtttgttt tttctttgt 1320
ccactatctt tctaagctc tctctctctc tgcatggctc ttggctgggt 1370

```

&lt;210&gt; 16

&lt;211&gt; 523

&lt;212&gt; DNA

WO 03/018768

PCT/US02/27144

&lt;213&gt; Canis familiaris

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 703528478J1

&lt;400&gt; 16

```
tgcatgagga ttccccaacc cagcccactg gtgttaatcc cctccttcc atgtttccac 60
tacaagggtat aaatacagcc cagagagtcc cgactgcagt tgatttcacc tgctttgtat 120
atagccatct ctacacattt ctgtacctct gcaagaacag gctccacaga ggtattcaaa 180
ataggtctgt acagaasaaa gcaaaagacat aaagcgtcac aagtggtaca aatcccgctc 240
atagcagcta tatactaacc cagcaaaaca gctttgcgca gtggcagtat cgtagccaat 300
gaggtttatc cagggtgaga ttattgctaa ttgaaaacta atccagcaaa acagagaaac 360
aattccaatc totgatttac atgttctccc tggcaattaa taatccagta acttctctag 420
ctatcttccc cataatgtct gccagcctt gttcctcacc ctgaacacta atttcgagat 480
cagactcaca cacagactag aaaaacaaca ggetgctcta tca 523
```

## 【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  
6 March 2003 (06.03.2003)

PCT

(10) International Publication Number  
WO 03/018768 A3

- (51) International Patent Classification: C07H 1/00, 5/04, 5/06, 19/00, 21/00, 21/02, 21/04, C08B 37/00, A01N 61/00, 37/18, 43/04, A61K 31/00, 31/70, 35/14, 38/00, 38/16, C12Q 1/00, 1/68, G01N 33/53, 33/537, 33/543, 33/567, 33/574, C12P 19/30, 19/34, 21/00, 21/08, C07K 1/00, 2/01, 4/00, 5/00, 7/00, 14/00, 16/00, 17/00, 17/14, G01N 33/68
- (21) International Application Number: PCT/US02/27144
- (22) International Filing Date: 22 August 2002 (22.08.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/314,914 24 August 2001 (24.08.2001) US
- (71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (72) Inventors; and  
(75) Inventors/Applicants (for US only): TRIBOULEY, Catherine, M. [FR/US]; 1121 Tennessee St. # 5, San Francisco, CA 94107 (US). LASEK, Amy, K, W. [US/US]; 297 41st St., Oakland, CA 94611 (US). YUE, Henry [US/US]; 826 14th Avenue, Sunnyvale, CA 94087 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US).
- (74) Agents: STREETER, David, G. et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (81) Designated States (national): AI, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GI, 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, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GI, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, UJ, UZ), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LI, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:  
with international search report
- (88) Date of publication of the international search report:  
14 August 2003
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 03/018768 A3

(54) Title: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER

(57) Abstract: The invention provides a transmembrane protein, TMDC, that is differentially expressed in a colon or stomach cancer. It also provides for the use of the protein, a cDNA encoding the protein, and antibodies that specifically bind the protein in various methods to diagnose, stage, treat, or monitor the progression or treatment of a colon or stomach cancer.

## 【 国際調査報告 】

INTERNATIONAL SEARCH REPORT		International application No. PCT/US02/27144		
<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC(7) : C07H 1/00, 5/04, 5/06, 19/00, 21/00, 21/02, 21/04; C08B 37/00; A01N 61/00, 37/18, 43/04; A61K 31/00, 31/70, 33/14, 38/00, 38/16; C12Q 1/00, 1/68; G01N 33/53, 33/537, 33/543, 33/567, 33/574; C12P 19/30, 19/34, 21/06, 21/08; C07K 1/00, 2/00, 4/00, 5/00, 7/00, 14/00, 16/00, 17/00, 17/14; C01N 33/48				
<b>B. FIELDS SEARCHED</b>				
Minimum documentation searched (classification system followed by classification symbols) U.S. : Please See Continuation Sheet				
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched				
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) GenCore database search of amino acid and nucleic acids				
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>				
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.		
X	WO 01/23417 A2 (SMITHKLINE BEECHAM BIOLOGICALS) 05 April 2001	1-70		
---	(05.04.2001), see entire document.	-----		
Y		71-73		
Y	AN et al. Synthesis of Novel 3'-C Methylene Thymidine and 5-Methyluridine/Cytidine H-Phosphonates and Phosphoramidites for New Backbone Modification of Oligonucleotides. J. Org. Chem. 20 April 2001, Vol. 66, pages 2789-2801, see entire document.	1-3 and 71-73		
X	Database GenCore, Accession Number AAF82460, nucleic acid database search, Accession number AAF82460, 29 June 2001 (29.06.2001), 2 sheets.	1-4		
X	Database GenCore, Accession Number AAF82460, nucleic acid and amino acid database search, Accession number AAF82460, 29 June 2001 (29.06.2001), 1 sheet.	1-4, 22 and 23		
X	Database GenCore, Accession Number AX108538, amino acid and nucleic acid database search, Accession number AX108538, 30 April 2001 (30.04.2001), 1 sheet.	1-4, 22 and 23		
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.				
* Special categories of cited documents: <table border="0" style="width: 100%;"> <tr> <td style="width: 50%;">               "A" document defining the general state of the art which is not considered to be of particular relevance                "B" earlier application or patent published on or after the international filing date                "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)                "O" document referring to an oral disclosure, use, exhibition or other means                "P" document published prior to the international filing date but later than the priority date claimed             </td> <td style="width: 50%;">               "*" 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                "&amp;" document member of the same patent family             </td> </tr> </table>			"A" document defining the general state of the art which is not considered to be of particular relevance "B" earlier application or patent published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"*" 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 "&" document member of the same patent family
"A" document defining the general state of the art which is not considered to be of particular relevance "B" earlier application or patent published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"*" 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 "&" document member of the same patent family			
Date of the actual completion of the international search 28 January 2003 (28.01.2003)		Date of mailing of the international search report 22 APR 2003		
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230		Authorized officer <i>Alana M. Harris</i> Alana M. Harris, Ph.D. Telephone No. (703)308-0196		

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

INTERNATIONAL SEARCH REPORT

PCT/US02/27144

Continuation of B. FIELDS SEARCHED Item 1:

536/1.11, 18.7, 22.1, 23.1, 23.4, 23.5, 24.1, 24.3, 24.5; 514/1, 2, 8, 43, 44, 45; 435/4, 6, 7.1, 7.21, 7.23, 7.92, 69.1, 89, 91.1, 91.2; 530/300, 350, 385, 386, 387.7, 388.1, 387.3, 388.1, 389.1, 391.1, 391.3, 391.5, 391.7; 436/64, 164

## 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)
A 6 1 K 31/7125	A 6 1 K 31/7125	4 B 0 6 5
A 6 1 K 38/00	A 6 1 K 39/395	D 4 C 0 8 4
A 6 1 K 39/395	A 6 1 K 39/395	N 4 C 0 8 5
A 6 1 K 45/00	A 6 1 K 39/395	V 4 C 0 8 6
A 6 1 K 48/00	A 6 1 K 45/00	4 H 0 4 5
A 6 1 P 35/00	A 6 1 K 48/00	
C 0 7 K 1/22	A 6 1 P 35/00	
C 0 7 K 14/82	C 0 7 K 1/22	
C 0 7 K 16/32	C 0 7 K 14/82	
C 1 2 M 1/00	C 0 7 K 16/32	
C 1 2 N 1/15	C 1 2 M 1/00	A
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 N 15/02	C 1 2 P 21/02	C
C 1 2 P 21/02	C 1 2 P 21/08	
C 1 2 P 21/08	C 1 2 Q 1/68	A
C 1 2 Q 1/68	G 0 1 N 33/15	Z
G 0 1 N 33/15	G 0 1 N 33/50	Z
G 0 1 N 33/50	G 0 1 N 33/53	M
G 0 1 N 33/53	G 0 1 N 33/574	A
G 0 1 N 33/574	G 0 1 N 37/00	1 0 2
G 0 1 N 37/00	C 1 2 N 5/00	A
	C 1 2 N 15/00	F
	C 1 2 N 15/00	C
	A 6 1 K 37/02	

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

(72) 発明者 ラセク、エイミー・ケイ・ダブリュ

アメリカ合衆国カリフォルニア州 9 4 6 1 1 ・ オークランド ・ # 5 ・ フォーティファースト 2 9 7

(72) 発明者 ユエ、ヘンリー

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

(72) 発明者 ボーグン、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 AA26 AA40 BB03 BB20 CB01 DA12 DA13 DA14 DA36 DA37  
 DA78 FB02 FB03  
 4B024 AA01 AA12 BA36 BA45 CA02 CA04 CA12 GA11 HA14 HA15  
 4B029 AA23 BB20 CC03  
 4B063 QA18 QA19 QQ02 QQ42 QQ52 QR56 QR59 QR62 QR77 QR80  
 QR82 QS24 QS25 QS33 QS34 QS39 QX01 QX07  
 4B064 AG01 AG27 AG31 CA10 CA19 CA20 CC24 CE13 DA14



专利名称(译)	跨膜蛋白在癌症中差异表达		
公开(公告)号	<a href="#">JP2005505267A</a>	公开(公告)日	2005-02-24
申请号	JP2003523619	申请日	2002-08-22
[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	トリボレーキャサリーンエム ラセクエイミーケイダブリュ ユエヘンリー ボーグンマライアール		
发明人	トリボレー、キャサリーン・エム ラセク、エイミー・ケイ・ダブリュ ユエ、ヘンリー ボーグン、マライア・アール		
IPC分类号	G01N33/50 A61K31/7088 A61K31/7105 A61K31/711 A61K31/7115 A61K31/7125 A61K38/00 A61K39/395 A61K45/00 A61K48/00 A61P35/00 C07K1/22 C07K14/705 C07K14/82 C07K16/32 C12M1/00 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N15/02 C12N15/09 C12P21/02 C12P21/08 C12Q1/68 G01N33/15 G01N33/53 G01N33/574 G01N37/00		
CPC分类号	C07K14/705 A61K38/00		
FI分类号	C12N15/00.ZNA.A A61K31/7088 A61K31/7105 A61K31/711 A61K31/7115 A61K31/7125 A61K39/395. D A61K39/395.N A61K39/395.V A61K45/00 A61K48/00 A61P35/00 C07K1/22 C07K14/82 C07K16/32 C12M1/00.A C12N1/15 C12N1/19 C12N1/21 C12P21/02.C C12P21/08 C12Q1/68.A G01N33/15.Z G01N33/50.Z G01N33/53.M G01N33/574.A G01N37/00.102 C12N5/00.A C12N15/00.F C12N15/00.C A61K37/02		
F-TERM分类号	2G045/AA26 2G045/AA40 2G045/BB03 2G045/BB20 2G045/CB01 2G045/DA12 2G045/DA13 2G045/DA14 2G045/DA36 2G045/DA37 2G045/DA78 2G045/FB02 2G045/FB03 4B024/AA01 4B024/AA12 4B024/BA36 4B024/BA45 4B024/CA02 4B024/CA04 4B024/CA12 4B024/GA11 4B024/HA14 4B024/HA15 4B029/AA23 4B029/BB20 4B029/CC03 4B063/QA18 4B063/QA19 4B063/QQ02 4B063/QQ42 4B063/QQ52 4B063/QR56 4B063/QR59 4B063/QR62 4B063/QR77 4B063/QR80 4B063/QR82 4B063/QS24 4B063/QS25 4B063/QS33 4B063/QS34 4B063/QS39 4B063/QX01 4B063/QX07 4B064/AG01 4B064/AG27 4B064/AG31 4B064/CA10 4B064/CA19 4B064/CA20 4B064/CC24 4B064/CE13 4B064/DA14 4B065/AA91X 4B065/AA92X 4B065/AA93Y 4B065/AC14 4B065/BA02 4B065/CA24 4B065/CA25 4B065/CA46 4C084/AA01 4C084/AA02 4C084/AA07 4C084/AA13 4C084/AA17 4C084/BA01 4C084/BA08 4C084/BA22 4C084/BA23 4C084/CA18 4C084/DA27 4C084/NA14 4C084/ZB26 4C085/AA13 4C085/AA14 4C085/AA33 4C085/CC03 4C085/DD62 4C086/AA01 4C086/AA02 4C086/AA03 4C086/BC42 4C086/BC43 4C086/CB07 4C086/DA38 4C086/EA17 4C086/EA18 4C086/MA01 4C086/MA04 4C086/NA14 4C086/ZB26 4H045/AA10 4H045/AA11 4H045/AA20 4H045/AA30 4H045/BA10 4H045/CA41 4H045/DA76 4H045/DA86 4H045/EA51 4H045/FA72 4H045/FA74 4H045/GA26		
优先权	60/314914 2001-08-24 US		
外部链接	<a href="#">Espacenet</a>		

#### 摘要(译)

本发明提供TMDC，其是在结肠癌或胃癌中差异表达的跨膜蛋白。此外，通过使用该蛋白质，编码该蛋白质的cDNA以及以各种方法特异性结合该蛋白质的抗体，诊断，分期或监测结肠直肠癌或胃癌的进展或治疗的方法同样。

(61) Int.Cl. <sup>7</sup>		F I	テーマコード (参考)
C 1 2 N 15/00		C 1 2 N 15/00 Z N A A	2 G O 4 5
A 6 1 K 31/7088		A 6 1 K 31/7088	4 B O 2 4
A 6 1 K 31/7105		A 6 1 K 31/7105	4 B O 2 9
A 6 1 K 31/7111		A 6 1 K 31/7111	4 B O 6 3
A 6 1 K 31/7115		A 6 1 K 31/7115	4 B O 6 4
		審査請求 未請求 予備審査請求 未請求	(全 141 頁) 最終頁に続く
(21) 出願番号	特願2003-523619 (P2003-523619)	(71) 出願人	301005050
(86) (22) 出願日	平成14年3月22日 (2002.3.22)		インサイト・ゲノミックス・インコーポレ イテッド
(85) 翻訳文提出日	平成16年2月23日 (2004.2.23)		アメリカ合衆国カリフォルニア州9430 4・パロアルト・ボートドライブ 31 60
(86) 国際出願番号	PCT/US2002/027144	(74) 代理人	100089266
(87) 国際公開番号	W02003/018768		弁理士 大島 陽一
(87) 国際公開日	平成15年3月6日 (2003.3.6)		トリボレー、キャサリーン・エム
(31) 優先権主張番号	60/314, 914	(72) 発明者	アメリカ合衆国カリフォルニア州9410 7・サンフランシスコ・#5・テネシース トリート 1121
(32) 優先日	平成13年3月24日 (2001.3.24)		
(33) 優先権主張国	米国 (US)		
最終頁に続く			