

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2004-521611

(P2004-521611A)

(43) 公表日 平成16年7月22日(2004.7.22)

(51) Int. Cl. ⁷	F I	テーマコード (参考)
C 1 2 N 15/09	C 1 2 N 15/00	Z N A A
A 6 1 K 38/00	A 6 1 P 1/04	2 G O 4 5
A 6 1 P 1/04	A 6 1 P 9/10	4 B O 2 4
A 6 1 P 9/10	A 6 1 P 11/06	4 B O 6 3
A 6 1 P 11/06	A 6 1 P 13/12	4 B O 6 4
	審査請求 未請求 予備審査請求 有	4 B O 6 5
		(全 85 頁) 最終頁に続く

(21) 出願番号	特願2002-524078 (P2002-524078)	(71) 出願人	301005050
(86) (22) 出願日	平成13年8月27日 (2001.8.27)		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15年2月24日 (2003.2.24)		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) 国際出願番号	PCT/US2001/026682	(74) 代理人	100089266
(87) 国際公開番号	W02002/018575		弁理士 大島 陽一
(87) 国際公開日	平成14年3月7日 (2002.3.7)	(72) 発明者	ウォーカー、マイケル・ジー
(31) 優先権主張番号	60/229, 253		アメリカ合衆国カリフォルニア州94089・サニーベイル・#80・ボレガスアベニュー 1050
(32) 優先日	平成12年8月30日 (2000.8.30)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 細胞周期で発現した遺伝子

(57) 【要約】

本発明は、細胞周期障害の診断及び治療の方法に用い得る、cDNA及びcDNAによりコード化されたタンパク質、及び抗体を提供する。

【特許請求の範囲】

【請求項 1】

SEQ ID NO : 1 - 10 またはその相補体の核酸配列を有する複数の cDNA を含む組成物。

【請求項 2】

核酸を含むサンプルの遺伝子発現を検出するために組成物を使用する方法であり、

(a) ハイブリダイゼーション化合物を形成する条件下で核酸に請求項 1 に記載の化合物をハイブリダイズする方法

(b) ハイブリダイゼーション化合物形成を検出する方法を含み、その方法では化合物形成はサンプル中の cDNA の遺伝子発現を示すことを特徴とする。

10

【請求項 3】

組成物の cDNA が基質に付着することを特徴とする請求項 2 に記載の方法。

【請求項 4】

遺伝子発現を標準と比較し、その遺伝子発現が細胞周期疾患を意味する場合の請求項 7 に記載の方法。

【請求項 5】

組成物を用いて複数の分子か化合物をスクリーニングする方法であり、

(a) 特異結合を可能とする条件下で、請求項 1 に記載の組成物と複数の分子あるいは化合物を混合する方法、

(b) 特異結合の検出および、それにより組成物の cDNA に特異結合する分子あるいは化合物を同定する方法を含む。

20

【請求項 6】

SEQ ID NO : 1、2、4 - 10 とその相補体から選択した核酸配列を含む cDNA 。

【請求項 7】

請求項 6 に記載の cDNA 及び標識成分または医薬用担体を含む組成物。

【請求項 8】

核酸を含むサンプルの発現を検出するために cDNA を使用する方法であり、

(a) ハイブリダイゼーション化合物をさらに形成する条件下で核酸に請求項 6 に記載の cDNA をハイブリダイズする方法、

(b) ハイブリダイゼーション化合物形成を検出する方法を含み、その方法では化合物形成はサンプル中の cDNA の発現を示すことを特徴とする。

30

【請求項 9】

組成物の cDNA が基質に付着することを特徴とする請求項 8 に記載の方法。

【請求項 10】

請求項 8 に記載の方法であり、発現を標準と比較し、その発現が細胞周期疾患を意味することを特徴とする。

【請求項 11】

複数の分子あるいは化合物をスクリーニングする cDNA を用いて、リガンドを同定、精製する方法であり、下記の a) - c) を含む。

40

(a) 特異結合を可能とする条件下で、請求項 6 に記載の cDNA と複数の分子あるいは化合物を混合する方法を含み、

(b) 結合した cDNA を回収する方法、

(c) cDNA をリガンドから解離して、精製したリガンドを得る方法を含む。

【請求項 12】

請求項 11 に記載の方法であり、複数の分子または化合物は DNA 分子、RNA 分子、ペプチド核酸、転写因子、エンハンサー、リプレッサー、擬態、タンパク質から選択される。

【請求項 13】

SEQ ID NO : 1、2、4 - 10 から選択した cDNA を有する発現ベクター。

50

【請求項 14】

請求項 13 に記載の発現ベクターを含む宿主細胞。

【請求項 15】

タンパク質を産生する cDNA の使用に関する方法であり、

(a) タンパク質発現のための条件下で請求項 14 に記載の宿主細胞を培養する方法、及び

(b) 細胞培地からタンパク質を回収する過程を含む。

【請求項 16】

請求項 15 に記載の方法で産生され、精製されたタンパク質あるいはその部分である。

【請求項 17】

請求項 15 に記載の方法で産生されたタンパク質と標識成分または医薬用担体を含む組成物。

10

【請求項 18】

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

(a) 特異結合を可能とする条件下で、請求項 16 に記載のタンパク質と複数の分子あるいは化合物を混合する方法を含み、

(b) 結合したタンパク質を回収する方法、

(c) タンパク質をリガンドから解離して、精製したリガンドを得る方法を含む。

20

【請求項 19】

請求項 18 に記載の方法であり、複数の分子を DNA 分子、RNA 分子、ペプチド核酸、ペプチド、擬態、タンパク質、アゴニスト、アンタゴニスト、抗体から選択することを特徴とする。

【請求項 20】

タンパク質を用いて抗体を調製及び精製する方法であり、

(a) 抗体反応を誘発する条件下で、請求項 16 に記載のタンパク質で動物を免疫化する方法、

(b) 動物抗体を単離する方法、

(c) 前記タンパク質を基質に付着する方法、

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

30

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

【発明の詳細な説明】

【0001】

(技術分野)

本発明は、既知の細胞周期遺伝子と同時発現により同定された cDNA、および細胞周期疾患の診断、予後、治療、治療法の評価におけるそれらの cDNA の使用に関連する。

【0002】

(発明の背景)

細胞分裂は、すべての生物の成長、修復、増殖にとって基本的なプロセスである。単細胞生物では各細胞分裂により生命体の数が 2 倍となる。また多細胞の種では、新たな生命体を作製するために、あるいは消耗あるいはプログラムした細胞死により消失した細胞を交換するために多数の回数の細胞分裂が必要とされる。細胞分裂周期の詳細は様々であるが、基本的なプロセスは 3 つの主要な現象からなる。最初の段階である間期には、細胞分裂、DNA の複製、本質的なタンパク質の作製が含まれる。2 番目の段階である有糸分裂では、核物質が分裂して細胞の両側に分離する。最終の段階である細胞質分裂は、細胞質の分裂である。細胞周期事象の順序およびタイミングは、種々のチェックポイントで正または負のメカニズムによるプロセスを制御する細胞周期調節システムの制御下にある。

40

【0003】

50

癌と免疫状態、疾病及び疾患は、正常な細胞増殖の調節障害と関連する。癌では、調節障害は、正常な細胞遺伝子の突然変異のイソ型である発癌遺伝子にしばしば起因する。これらの発癌遺伝子の中には、ウイルスのゲノムの宿主細胞のDNAへの組み込みの結果としてウイルスにより活性化する場合もある。感染した細胞を連続的に細胞分裂をさせる能力のある2つ以上の発癌遺伝子が活性化することもある。他の発癌遺伝子は、発現の位置またはレベルに関して異常に発現する。後者のカテゴリーは、細胞増殖の転写調節を変更することにより癌の原因となる。少なくとも5つの発癌遺伝子の分類が知られている。分類にはサイトカイン、成長因子、*erbA*、*erbB*、*neu*、*ros*等の受容体、*src*、*yes*、*fps*、*abl*、*met*等の細胞内シグナルトランスデューサ、*fos*などの核転写因子、*Rb*、*p53*等の細胞周期制御タンパク質、*mdm2*、*sec*、*ras*等の変異腫瘍抑制遺伝子がある(Bohmannら(1987) *Science* 238: 1386-1392; CohenおよびCurran(1988) *Mol Cell Biol* 8: 2063-2069、van Straatenら(1983) *Proc Natl Acad Sci* 80: 3183-3187)。

10

20

【0004】

例えば癌においては、発癌遺伝子は成長因子シグナルの受容と伝達、そしてこれらのシグナルに反応する遺伝子発現の調節に関与することにより抑制のない細胞増殖に寄与する。成長因子による細胞の刺激は2セットの遺伝子すなわち初期反応遺伝子と遅延応答遺伝子を活性化させる。*myc*、*fos*、*jun*のプロトオンコジーンを含む初期反応遺伝子は、すべて遺伝子調節タンパク質をコード化する。これらの調節タンパク質は、細胞周期進行に関与するサイクリン、サイクリン依存キナーゼなどのタンパク質をコード化する遅延反応遺伝子の転写を活性化する。

【0005】

既知の細胞周期遺伝子と同時発現するcDNAの発見により、細胞周期疾患の診断、予後、治療、治療の評価において有用な新規の組成物を提供して当分野における必要を満たす。

【0006】

(発明の概要)

本発明は、複数の生物サンプルの1つ以上の既知の細胞周期遺伝子と同時発現するSEQ ID NO: 1-10またはその相補体の核酸配列を有する複数のcDNAを含む組成物を提供する。本発明はさらに、前記の組成物の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ために、複数の分子をスクリーニングする目的である組成物を使用する方法、そして特異結合を可能にする条件下で分子と前記の組成物を混合し、そして特異結合を検出してcDNAに特異結合するリガンドを同定する方法を提供する。ある実施形態では、それらの分子はDNA分子、RNA分子、ペプチド核酸、転写因子、エンハンサー、リプレッサー、擬態、タンパク質から選択される。

30

【0007】

本発明は、核酸を含むサンプルの遺伝子発現を検出するためにある組成物を使用する方法を提供し、その方法には、1つ以上のハイブリダイゼーション複合体を形成する条件下で核酸にその組成物をハイブリダイズさせる方法、そして複合体形成がサンプル中の遺伝子発現を示すようにしてハイブリダイゼーション複合体形成を検出する方法が含まれる。一実施例では、前記の組成物中のcDNAをある基質に付着させる。別の実施例では、標準に比較した時の複合体形成が細胞周期疾患の診断となる。

40

【0008】

本発明は、SEQ ID NO: 1、2と4-10及びその相補体から選択する核酸配列を有する単離したcDNAを提供する。異なる実施形態では、各cDNAを診断用として、プローブとして、発現ベクターで、および細胞周期疾患の予後と治療の評価において使用する。さらに本発明は、cDNA及び標識成分を含む組成物を提供する。本発明はさらに、前記のcDNAに特異結合するリガンドを同定するために、複数の分子をスクリーニングする目的で前記のcDNAを使用する方法、そして特異結合を可能にする条件下でサン

50

ブルと前記の cDNA を混合させ、結合した cDNA を回収し、そして結合した cDNA からリガンドを分離し精製したリガンドを獲得する方法を含めた方法を提供する。ある実施形態では、それらの分子は DNA 分子、RNA 分子、ペプチド核酸、転写因子、エンハンサー、リプレッサー、擬態、タンパク質からスクリーニングして選択される。

【0009】

本発明は、核酸を含むサンプルの遺伝子発現を検出するために cDNA を使用する方法を提供する。また 1 つ以上のハイブリダイゼーション複合体を形成する条件下でサンプルの核酸に cDNA をハイブリダイズする、そして複合体形成がサンプル中の遺伝子発現を示すハイブリダイゼーション複合体形成を検出する方法を含む。一実施例では、cDNA を基質に付着する。別の実施例では、標準と比較した時の遺伝子発現が細胞周期疾患の診断となる。更に該方法により、cDNA を含有するベクター、およびベクターを含有する宿主細胞を提供する。また cDNA によりコード化されるタンパク質やペプチドを産生する宿主細胞を使用する方法を提供し、その内には、タンパク質の発現用の条件下で宿主細胞を培養する方法と細胞培養からタンパク質を回収する方法が含まれる。本発明は、本発明の cDNA によりコード化される精製したタンパク質を提供する。本発明はさらに、タンパク質を特異結合するリガンドを同定、精製する複数の分子をスクリーニングするためにタンパク質やペプチドを使用する方法を提供する。ある実施形態では、分子は DNA 分子、RNA 分子、ペプチド核酸、タンパク質、アゴニスト、アンタゴニスト、抗体からスクリーニングして選択される。

10

【0010】

さらに本発明は、抗体を調製かつ精製するために、あるタンパク質を使用する方法を提供し、その方法の内には抗体反応を誘発する条件下でタンパク質またはペプチドによって動物を免疫化すること、動物抗体を単離すること、ある基質にタンパク質を付着させること、そのタンパク質への特異結合を可能にする条件下でその基質を単離した抗体と接触させること、タンパク質から抗体を解離して精製した抗体を得ることが含まれる。本発明はまた、細胞周期疾患を診断するためにタンパク質を特異結合する抗体を使用する方法、そして特異結合の条件下でサンプルに抗体を混合させ、抗体複合物形成を検出し、標準と抗体複合物形成を比較し、それによって細胞周期疾患を診断する方法を提供する。さらに本発明は、細胞周期疾患を治療で使用する cDNA、タンパク質、またはタンパク質あるいはペプチドを特異結合する抗体そして医薬用担体を含む組成物を提供する。

20

30

【0011】

(発明を実施するための形態)

請求の範囲及び明細書中で用いている単数形の「或る」及び「その(この)」の表記は、文脈から明らかにそうでないとされる場合を除いて複数のものを指す場合もあることに注意しなければならない。従って、例えば「或る宿主細胞」と記されている場合にはそのような宿主細胞が複数あることもあり、「或る抗体」と記されている場合には単数または複数の抗体、及び、当業者に公知の抗体の等価物等についても言及しているのである。

【0012】

(定義)

「アレイ」は、基質上の少なくとも 2 つの cDNA あるいは抗体の規則正しい配列を指す。cDNA あるいは抗体の少なくとも 1 つが調節または標準を表す。そして他方が目的の診断または薬剤の cDNA あるいは抗体を表す。基質上の 2 ~ 約 40,000 の cDNA あるいは 2 ~ 約 40,000 のモノクローナル抗体またはポリクローナル抗体の配列では、各 cDNA と少なくとも 1 つの核酸の間で形成される各標識ハイブリダイゼーション複合体のサイズとシグナル強度、あるいは各抗体と抗体が特異結合する少なくとも 1 つのタンパク質の間で形成される抗体タンパク質複合体は、確実に個別に区別できる。

40

【0013】

「細胞周期遺伝子」は、制御されていない細胞周期と関連する診断、予後、治療、治療の評価において有用であると前もって同定された cDNA を指す。通常は、これは任意の組織で正常な発現と比較する時に細胞周期疾患を患う患者の組織でより高い(低い)レベル

50

で既知の遺伝子が示差的に発現されていることを意味する。本発明で使用され、実施例4で記載される細胞周期遺伝子は、*c d c 2*、*c d c 7*、*c d c 2 3*、サイクリンB、*h B u b 1*、*H K S P*、*h p 5 5 c d c*、*M C A K*、*m i t o s i n*、*m k i 6 7 a*、*M K L P - 1*、*m y b*、*n l k 1*、*c d c 2 1*、*P R C 1*、*A i k 2*、サバイピン、*t o p o I I*、*U b c H 1 0*がある。

【0014】

「細胞周期疾患」とはすべての癌または免疫疾患を指す。限定するものではないが、腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫または血、骨、骨髄、脳、胸、胃腸管（食道、胃、小腸または結腸）、心臓、腎臓、肝臓、肺、リンパ、筋肉、神経、卵巣、膵臓、前立腺、皮膚、脾臓、精巣と子宮の癌、喘息、アテローム性動脈硬化症、クローン病、糸球体腎炎、多発性硬化症、重症筋無力症、骨粗鬆症、慢性関節リウマチ、硬皮症、そして、全身エリテマトーデスが含まれる。

10

【0015】

「*c D N A*」とは、単離したポリヌクレオチド、あるいはそのポリヌクレオチドの任意の断片またはオリゴヌクレオチドを指す。それは、ゲノム起源または合成起源、二本鎖または一本鎖であってもよく、特別の作用を起こすため、あるいは有用な組成物形成するために、糖質、脂質、タンパク質その他の物質と結合したものでよい。

【0016】

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

20

【0017】

「分離したあるいは精製した」とは、自然環境から分離され、かつ自然に存在する他の化合物から分離している*c D N A*あるいはタンパク質を指す。

【0018】

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

【0019】

用語「タンパク質」は、天然または合成のペプチド或いはそれらの任意の断片またはオリゴペプチドを指す。

30

「サンプル」は、核酸、タンパク質、抗体その他を含むようなその最も広い意味で用いられる。サンプルは体液、細胞調製の可溶性分画、細胞が成長する培地のアリコット、染色体、細胞小器官、あるいは細胞から単離または抽出された膜、溶液中のまたは基板に固定されたゲノムDNA、RNA、または*c D N A*と、細胞、組織、組織プリント、フィンガープリント、口内細胞、皮膚あるいは髪などから成る可能性がある。

【0020】

「類似性」とは、*S m i t h - W a t e r m a n*アラインメント (*S m i t h*及び*W a t e r m a n (1 9 8 1) J M o l B i o l 1 4 7 : 1 9 5 - 1 9 7*)あるいは*B L A S T 2 (A l t s c h u l r a (1 9 9 7) N u c l e i c A c i d s R e s 2 5 : 3 3 8 9 - 3 4 0 2*)などの標準化されたアルゴリズムを用いてアラインメントされた少なくとも2つの配列間で一致するヌクレオチドあるいは残基の定量化（通常は%）を指す。*B L A S T 2*は、再現性のある方法で、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か、または2つの配列をより有意に比較するために使用し得る。タンパク質では特に、保存的置換では類似性が同一性よりも大きい。例えば、ロイシンまたはイソロイシンのバリンは報告された割合を計算する時にカウントされる。保存と考えられた置換が当分野で公知である。

40

【0021】

「特異結合」とは、構造、特に分子側基に依存する、2つの分子間での特別で精確な相互

50

作用を指す。例えば、調節タンパク質のDNA分子の主溝への挿入、あるいはタンパク質のエピトープとアゴニスト、アンタゴニストまたは抗体との間の結合がある。

【0022】

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

【0023】

「転写イメージ」とは、特定の時に特定の組織の遺伝子転写活性のプロフィールである。

【0024】

「変異体」とは、cDNAまたはcDNAがコードするタンパク質の認識される変異である分子を指す。スプライス変異体は、スコアが少なくとも100、最も好ましいのは少なくとも400である、BLASTスコアにより決定する可能性がある。対立遺伝子変異体はcDNAに対して高一致率を有し、100塩基に付き約3塩基が異なる。「1塩基多型性」(SNP)とは、欠失、挿入または置換による単一の塩基での変異を指す。変異は保存(プリンからプリン)あるいは非保存(プリンからピリミジン)される可能性があり、コード化されたアミノ酸または2次、3次、4次構造の変異となる可能性もならない可能性もある。

【0025】

(発明)

本発明は、特定の疾患、調節経路、細胞内部分、細胞の種類、組織の種類あるいは種と関連するcDNAまたはタンパク質を同定する方法を利用する。特にその方法は細胞周期疾患のための治療の診断、予後、治療、評価において有用なcDNAを同定する。

【0026】

その方法は、複数のライブラリで発現するcDNAの同定法を提供する。既知の機能の遺伝子の発現パターンを未知の機能のcDNAのそれと比較することにより、特異な同時発現確率閾値を満たすかどうかを決定する。この比較により、既知の遺伝子と同時発現確率の高いcDNAのサブセットを同定することが可能となる。

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

【0027】

本発明の同時発現分析で使用されるcDNAライブラリは、副腎、胆道、膀胱、血球、血管、骨髄、脳、気管、軟骨、クロム親和系、結腸、結合組織、培養細胞、胚幹細胞、内分泌腺、上皮組織、食道、胎児、神経節、心臓、視床下部、免疫機構、腸、ランゲルハンス島、腎臓、咽頭、肝臓、肺、リンパ、筋肉、神経、卵巣、膵臓、陰茎、末梢神経系、腹膜、食細胞、下垂体、胎盤、胸膜、前立腺、唾液腺、精嚢、骨格、脾臓、胃、精巣、胸腺、舌、尿管、子宮等から得ることが可能である。選択されたcDNAライブラリの数の範囲は、わずか5から10,000以上まで可能性がある。好ましくはcDNAライブラリの数は500以上である。

【0028】

好適な実施例では、cDNAを単一転写に由来する配列断片などの関連配列から構築する。ポリヌクレオチドの構築は、限定するものではないがEST、ESTの伸長、クローン化インサートのショットガン・シーケンシングまたは全長cDNA等多様な種類の配列を使用して行うことが可能である。最も好適な実施例では、1999年3月25日に提出したUS 9,276,534で開示したアルゴリズムを使用して構築したヒト配列

10

20

30

40

50

から cDNA を導出する。該特許は引用することをもって本明細書の一部とする。

【0029】

実験的に、限定するわけではないが、空間的固定またはゲル電気泳動法によるディファレンシャルディスプレイ、ゲノム不一致スキニング、RDA (representational difference analysis) 法、転写イメージング等の方法により cDNA の示差発現を評価することができる。SEQ ID NO: 1、5、10 の代表的転写イメージは実施例 15 に見出される。転写イメージは、本明細書で開示する同時発現方法により作成したデータを確認する。さらに示差発現をマイクロアレイ技術により評価することができる。これらのいずれの方法を単独あるいは併用して使用することができる。

10

【0030】

制御不能な細胞増殖と関連のある病気の診断あるいは予後マーカーとしてあるいは治療標的としての遺伝子の機能と使用に基づいて、既知の細胞周期遺伝子を選択することができる。好ましくは既知の細胞周期遺伝子は、cdc2、cdc7、cdc23、サイクリン B、hBub1、HKSP、hp55cdc、MCAK、mitosin、mki67a、MKLP-1、myb、nlk1、cdc21、PRC1、Aik2、サバイピン、topoII、UbcH10 がある。

【0031】

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

20

【0032】

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

30

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

10

【0033】

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

20

【0034】

所定の対の遺伝子には、出現データを2 x 2 分割表(下記)で概略する。

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

30

【0035】

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

40

50

常有意と考慮される（前出の *Agrest i* と *Rice*）。

【0036】

2つの遺伝子の同時発現の確率を評価する方法には幾つかの前提がある。方法はライブラリが独立しており、同様にサンプル化していると想定する。しかし実際の状況では、選択したcDNAライブラリは完全には独立していない。1つ以上のライブラリを単一の対象または組織から得ることが可能だからである。また完全には同様にサンプル化されてもいない。異なる数のcDNAを各ライブラリから配列化する可能性があるからである。通常シーケンスしたcDNAの数の範囲は、ライブラリ毎に5,000から10,000cDNAである。更に、フィッシャーの正確同時発現確率を各遺伝子対少なくとも5つのライブラリで出現する他のすべての構築した37,071の遺伝子に対して計算する。また複数の統計試験のためにボンフェローニ検定を用いた。 10

【0037】

上記の方法を用いて、細胞周期と特異的である既知の遺伝子と強い関連または同時発現を示したcDNAを同定した。実施例5に見られる同時発現表に記載される結果は、下記の表に要約されている。列1はSEQ ID番号、列2はcDNAが最も高度に同時発現する既知の細胞周期遺伝子、列3はp値、列4は同時発現したcDNAが特異的診断マーカーである細胞周期疾患である。

SEQ ID	細胞周期遺伝子	p値	細胞周期疾患	
1	topo II	16	膵臓神経内分泌カルチノイド	20
2	PRC1	12	結腸癌	
3	CDC23	12	リンパ腫	
4	topo II、PRC1	10	転移性黒色腫	
5	サイクリン B、 UbcH10	13	乳癌	
6	PRC1	16	結腸癌	30
7	サイクリンB	9.5	脳腫瘍	
8	topo II	13	精巣癌	
9	topo II	9	転移性黒色腫	
10	hp55cdc	17	結腸癌	

【0038】

この表は本明細書で請求しているcDNAが既知の細胞周期遺伝子との間で、著しく高い同時発現（00000001未満）を有することを示す。よって、細胞周期疾患の診断、予後、治療の評価においてcDNAは代理マーカーとして有用であり、制御不能な細胞周期の排除と制御のための医薬として役立つ可能性がある。さらにcDNAから発現したタンパク質またはペプチドは、治療に役立つ可能性がある、あるいは同定や発達のための標的を有する。同様に、タンパク質から作製されたあるいはタンパク質を使用して同定した抗体は、治療用や薬理用のキャリアーに役立つ可能性がある。 40

【0039】

よって本発明は、一実施例ではSEQ ID NO: 1-10またはその相補体の核酸配列を有するcDNAの組成物を含む。これらの10のcDNAは、既知の細胞周期遺伝子と 50

の強い同時発現、また互いに強い同時発現を有することが、本発明の方法により示される。さらに本発明は、cDNA、相補体、SEQ ID NO: 1、2、4 - 10から選択したcDNAを含むプローブを提供する。変異体は通常はこれらの配列の少なくとも1つに対して少なくとも約70%の核酸配列同一性を有する。

【0040】

cDNAまたはコード化したタンパク質を、以前に同定してアノテーションが付けられたモチーフ、配列、遺伝子機能を含む、GenBankの霊長類 (pri)、げっ歯類 (rod)、哺乳動物 (mam)、脊椎動物 (vrt p)、真核生物 (euk p) のデータベースと、SwissProt、BLOCKS (Bairoch ら (1997) *Nucleic Acids Res* 25:217-221)、PFAMその他のデータベース 10
 に対して検索するために使用することができる。Basic Local Alignment Search Tool (BLAST; Altschul (1993) *J Mol Evol* 36:290-300、Altschul ら (1990) *J Mol Biol* 215:403-410)、BLOCKS (Henikoff及びHenikoff (1991) *Nucleic Acids Res* 19:6565-6572)、隠れマルコフモデル (HMM; Eddy (1996) *Curr Opin Str Biol* 6:361-365、Sonnhammer ら (1997) *Proteins* 28:405420) 等のアルゴリズムの他に二次構造ギャップペナルティを有する一次配列パターン (Smith ら (1992) *Protein Engineering* 5:35-51) を検索する方法をヌクレオチドとアミノ酸配列を操作、分析するために使用する 20
 ことができる。これらのデータベース、アルゴリズム、またその他の方法は当分野で公知であり、Ausubel ら (1997; Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7) と Meyers (1995; Molecular Biology and Biotechnology, Wiley VCH, New York NY, 856-853ページ) により記載されている。

【0041】

また本発明には、ストリンジェント条件下で、SEQ ID NO: 1 - 10とその断片にハイブリダイズすることができるポリヌクレオチドが含まれている。塩分濃度、温度及びその他の化合物、当分野で知られている条件によりストリンジェント条件を定義すること 30
 ができる。例えばプレハイブリタイゼーション、ハイブリダイゼーション、洗浄溶液中の塩分の濃度を変えることにより、あるいはハイブリダイゼーションと洗浄温度を変えることにより条件を選ぶことができる。幾つかの基質で、ホルムアミドをプレハイブリタイゼーションとハイブリダイゼーション溶液に加えることにより温度を下げる 40
 ことができる。

【0042】

ハイブリダイゼーションは、1%のドデシル硫酸ナトリウム (SDS) を 60 で 5 x SSC (クエン酸ナトリウム食塩水) などのバッファーで低ストリンジェンシーで実施することができる。そして不一致を含む核酸配列の間で化合物の形成を可能にする。続いて、完全な相補的な配列を含有するそれらの化合物のみのハイブリダイゼーションを維持するために 0.1% SDS を 45 (中間ストリンジェンシー) または 68 (高ストリン 40
 ジェンシー) で、0.2 x SSC などのバッファーで高いストリンジェンシーで洗浄を実施した。バックグラウンドシグナルは SDS、Sarkosyl または TRITON X-100 (Sigma-Aldrich, St. Louis MO) などの界面活性剤または変性したサケ精子DNAなどのブロッカーの使用により減少させることができる。ハイブリダイゼーション方法については Ausubel (前出 units 2.8 - 2.11、3.18 - 3.19、4-6-4.9)、Sambrook ら (1989; Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY) に詳述されている。

【0043】

10

20

30

40

50

cDNAを、部分的ヌクレオチド配列を利用し且つ当分野でよく知られているPCR法をベースにした種々の方法を用いて伸長させ、プロモーター及び調節要素等の上流配列を検出することができる。(Dieffenbach及びDveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview NYを参照)。さらにXL-PCRキット(Applied Biosystems (ABI), Foster City CA)、ネステッドプライマー、市販のcDNAライブラリ(Life Technologies, Rockville MD)あるいはゲノムライブラリ(Clontech, Palo Alto CA)を使用して、配列を伸長することができる。全てのPCRベースの方法に対して、市販されているソフトウェア(LASERGENE software, DNASTAR, Madison WI)或いは別のプログラムを用いて、長さが約15~30ヌクレオチド、GC含有率が約50%以上、温度約68~72でハイブリダイゼーション複合体を形成するようにプライマーを設計し得る。

【0044】

本発明の別の実施例では、宿主細胞内でタンパク質またはその構造的また機能的部分の発現を誘導する組換えベクターにこのcDNAをクローニングすることが可能である。遺伝暗号固有の縮重に起因して、同一或いは機能的等価のアミノ酸配列をコードするような別のDNA配列を作製して、このcDNAによりコードするタンパク質の発現に利用し得る。種々の目的でヌクレオチド配列を変えるために、本発明のヌクレオチド配列を当分野で通常知られている方法を用いて組み換えることができる。ここで目的には、限定するものではないが遺伝子産物のクローニング、プロセッシング、発現の調節がある。遺伝子断片及び合成オリゴヌクレオチドのランダムなフラグメンテーション及びPCR再アセンブリによるDNAシャッフリングを用い、ヌクレオチド配列を組み換えることが可能である。例えば、オリゴヌクレオチドを介した部位特異的変異誘導を利用して、新規な制限部位の作製、グリコシル化パターンの変更、コドン優先の変更、スプライス変異体の生成等を起こす突然変異を導入し得る。

【0045】

生物学的に活性なタンパク質を発現させるために、cDNAまたはその誘導体を発現ベクターに挿入することができる。発現ベクターとは即ち特定の宿主内で挿入されたコーディング配列の転写及び翻訳の調節に必要な要素を含むベクターである。これらの要素には、エンハンサー、構成型及び発現誘導型プロモーター、5'及び3'の非翻訳領域などの調節配列がある。当業者に周知の方法を用いて、発現ベクターを作製することが可能である。これらの方法には、in vitro組換えDNA技術、合成技術、及びin vivo遺伝子組換え技術が含まれる(前出のSambrook、前出のAusubel)。

【0046】

多様な発現ベクター/宿主細胞系を用いてcDNAを発現することができる。限定するものではないがこのような宿主系には、組換えバクテリオファージ、プラスミドまたはコスミド発現ベクターで形質転換させた細菌や、酵母菌発現ベクターで形質転換させた酵母菌や、バキュロウイルス発現ベクターで感染させた昆虫細胞系や、ウイルスあるいは細菌性発現ベクターを含む発現ベクターで形質転換させた植物細胞系、あるいは動物細胞系などの微生物がある。長期にわたり哺乳動物系内で組換えタンパク質を生成するためには、細胞株内の安定発現が望ましい。例えば、発現ベクターであって複製発現因子、内在性発現因子、或いはその両者のウイルス起源を含むものと、同一或いは別のベクター上の選択可能または可視マーカー遺伝子とを用いて、cDNAを細胞株に形質転換することが可能である。本発明は使用されるベクターまたは宿主細胞によって限定されるものではない。

【0047】

一般に、cDNAを含み且つタンパク質を発現する宿主細胞は、当業者によく知られている種々の方法を用いて同定することが可能である。限定するものではないが当業者によく知られている方法には、DNA-DNA或いはDNA-RNAハイブリダイゼーション、PCR法、核酸或いはアミノ酸配列の検出、定量、或いはその両方を行うための膜系、溶

10

20

30

40

50

液ベース或いはチップベースの技術を含むタンパク質バイオアッセイまたはイムノアッセイ技術がある。特異的なポリクローナル抗体またはモノクローナル抗体のどちらかを用いるタンパク質の発現の検出及び計測のための免疫学的な方法は、当分野で周知である。このような技法には、酵素に結合した免疫吸着剤検定法（E L I S A）、ラジオイムノアッセイ（R I A）、フローサイトメーター（F A C S）などがある。

【0048】

c D N A で形質転換された宿主細胞は、細胞培地でのこのタンパク質の発現及び回収に好適な条件下で培養される。遺伝形質転換細胞から製造されたタンパク質が分泌されるか細胞内に留まるかは、使用される配列、ベクター、或いはその両者に依存する。c D N A を含む発現ベクターは、原核細胞膜及び真核細胞膜を透過するタンパク質の分泌を誘導するシグナル配列を含むように設計できることは、当業者には理解されよう。

10

【0049】

更に、宿主細胞株の選択は、挿入した配列の発現を調節する能力または発現したタンパク質を所望の形に処理する能力によって行い得る。限定するものではないがこのようなタンパク質の修飾には、アセチル化、カルボキシル化、グリコシル化、リン酸化、脂質化及びアシル化がある。タンパク質の「プレプロ」形を切断するような翻訳後処理を利用して、タンパク質のターゲティング、折りたたみ及び/または活性を特定することも可能である。翻訳後の活性のための固有の細胞装置及び特徴のある機構を有する種々の宿主細胞（例えば C H O、H e L a、M D C K、M E K 2 9 3、W I 3 8 等）は、A T C C（M a n a s s a s, V A）から入手可能であり、発現タンパク質の正しい修飾及び処理を確実に

20

【0050】

本発明の別の実施例では、自然或いは変更された、または組換えの核酸配列を上記した任意の宿主系の異種タンパク質成分を含む融合タンパク質の翻訳となる異種配列に結合させる。異種タンパク質部分は市販されている親和性基質を用いて融合タンパク質の精製を促進し得る。限定されるものではないがこのような部分には、グルタチオンSトランスフェラーゼ（G S T）、マルトース結合タンパク質、チオレドキシン、カルモジュリン結合ペプチド、6 - H i s、F L A G、c - m y c、赤血球凝集素、モノクローナル抗体エピトープがある。

【0051】

別の実施例によれば、c D N A は、当分野で周知の化学的または酵素的方法を用いて、全体或いは一部が合成可能である（例えば、C a r u t h e r s . . ら（1980）N u c l . A c i d s S y m p . S e r 7 : 2 1 5 - 2 3 3、前出 A u s u b e l を参照）。例えば、種々の固相技術（R o b e r g e ら（1995）S c i e n c e 2 6 9 : 2 0 2 2 0 4）を用いてペプチド合成を行うことができる。また A B I 4 3 1 A ペプチドシンセサイザ（A B I）等の装置を用いて合成を自動化することができる。所望であれば、変異体を産生するために、合成中に及び/または他のタンパク質からの配列との組み合わせでアミノ酸配列を変更することができる。

30

【0052】

スクリーニング、診断及び治療

組成物または c D N A を細胞周期疾患の診断、予後、治療、治療法の選択と評価に用いることができる。細胞周期疾患とは限定するものではないが、腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫または血、骨、骨髄、脳、胸、胃腸管（食道、胃、小腸または結腸）、心臓、腎臓、肝臓、肺、リンパ、筋肉、神経、卵巣、膵臓、前立腺、皮膚、脾臓、精巣と子宮の癌、喘息、アテローム性動脈硬化症、クローン病、糸球体腎炎、多発性硬化症、重症筋無力症、骨粗鬆症、慢性関節リウマチ、硬皮症、そして全身エリテマトーデスが含まれる。

40

【0053】

組成物または c D N A を使用して、特異結合親和性のために複数の分子をスクリーニングする。このアッセイを、生物系におけるポリヌクレオチドの活性を調節する複数の D N A

50

分子、RNA分子、ペプチド核酸(PNA)、ペプチド、リボザイム、抗体、アゴニスト、アンタゴニスト、抗体、免疫グロブリン、インヒビター、そして転写因子、エンハンサー、リプレッサー、薬剤その他などのタンパク質をスクリーニングするために使用することができる。このアッセイには、複数の分子を提供すること、特異結合を可能とするのに好適な条件下でcDNAまたはその断片を複数の分子と混合すること、cDNAを特異結合する少なくとも1つの分子を同定する特異結合を検出することが含まれる。

【0054】

同様に、タンパク質またはその部分を多様なスクリーニングアッセイの任意の分子または化合物のライブラリをスクリーニングするために用い得る。そのようなスクリーニングで用いられたタンパク質の部分は、溶液中で遊離するか、非生物的基質または生物的基質に固定させるか(例えば細胞表面上に保持される)、あるいは細胞内に位置することになる。タンパク質と分子間の特異結合を測定することは可能である。このタンパク質に特異結合する複数のDNA分子、RNA分子、PNA、ペプチド、様態、リボザイム、抗体、アゴニスト、アンタゴニスト、免疫グロブリン、インヒビター、ペプチド、ポリペプチド、薬剤その他などのタンパク質をスクリーニングするためにこのアッセイを使用することができる。非常に小さなアッセイ量と微量の試験化合物を使用して高い処理能力でスクリーニングするための方法がBurbbaum らのUSPN 5,876,946に記載されている。当該文献は特に引用することを以って本明細書の一部となす。また当方法により酵素阻害または受容体結合のための多数の分子をスクリーニングする。

10

【0055】

好適な一実施例では、不在、存在または遺伝子の正常な標準発現と比較して変更 増加あるいは減少を決定するために診断目的でcDNAを使用する。ポリヌクレオチドは、相補的RNAおよびDNA分子、分枝核酸またはPNAから成る。別の実施例では、cDNAの発現が疾患と相関するサンプル中の遺伝子発現を検出及び定量するためにcDNAを使用する。別法では、疾患と関連する遺伝多型性を検出するcDNAを使用することができる。これらの多型性を転写cDNA中で検出することができる。

20

【0056】

固有の領域、調節領域または保存されたモチーフに由来するかに応じてプローブの特異性を決定する。プローブが同定するのは、天然に存在する正確に相補的な配列、対立遺伝子変異体または関連配列のみであるか否かは、プローブの特異性及び診断ハイブリダイゼーション或いは増幅(極大、高、中間あるいは低)のストリンジェンシーの両方によって決定されることになる。関連する配列を検出するよう設計したプローブは、それらのcDNAのいずれに対して好ましくは少なくとも50%の配列同一性を有するべきである。

30

【0057】

ハイブリダイゼーションプローブの作製方法には、mRNAプローブの作製のためのベクターに核酸配列をクローニングする方法がある。ベクターは、当業者に知られており、市販されており、RNAポリメラーゼ及び標識されたヌクレオチドを加えることによって、in vitroでRNAプローブを合成するために用いられ得る。ハイブリダイゼーションプローブは、種々のレポーターの集団によって標識され得るヌクレオチドを組み込むことが可能である。限定するわけではないが、レポーター集団の例としては、³²Pまたは³⁵S等の放射性核種、或いはアビジン/ビオチン結合系を介してプローブに結合されたアルカリホスファターゼ等の酵素標識、蛍光標識その他が挙げられる。標識化cDNAは、変異タンパク質発現を検出するために被験者からのサンプルを使用するサザン法やノーザン法、ドットプロット法、或いはその他の膜系の技術、PCR法、マイクロアレイに使用することが可能である

40

【0058】

cDNAを標準的な方法で標識化して、ハイブリダイゼーション複合体の形成と検出に好適な条件下で患者のサンプルに加える。インキュベーション後、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シグナルの量を定量して標準値と比較する。標準値は、疾患の疑いのない任意の対照サンプルから通常は得る。患者サンプルのシグナル量が

50

対照サンプルと比べて著しく変化している場合は、サンプル内の発現の変異レベルの存在は疾患の存在を示している。前もって確立された標準と患者サンプル中に形成されたハイブリダイゼーション複合体を比較するための定性または定量方法は、当分野で公知である。

【0059】

このようなアッセイは、動物実験、臨床試験における特定の治療効果を推定するため、或いは個々の患者の治療をモニターするために用いることもできる。症状の存在を一旦確定して治療プロトコルを開始すると、ハイブリダイゼーションまたは増幅アッセイを通常ペースで繰り返して、患者の発現レベルが正常な被検者に観察されるレベルに近づき始めたかどうかを測定する。連続アッセイから得られた結果を用いて、数日から数年の期間にわたる治療の効果を示し得る。

10

【0060】

発現パターンをモニターするために、cDNAをマイクロアレイ上で使用することも可能である。マイクロアレイはまた、スプライス変異体、突然変異及び多型性の同定に用いることができる。発現パターンの分析から得た情報を用いることで、遺伝子機能を決定し、疾患の遺伝的根拠を理解し、疾患を診断し、疾病治療における薬剤の活性を開発及びモニターすることができる。マイクロアレイはまた、ゲノムレベルで特定の個体群を特徴付ける遺伝多様性、1塩基多型性の検出に用いることができる。

【0061】

しかし別法では、cDNAを用いて、天然のゲノム配列をマッピングするのに有用なハイブリダイゼーションプローブを作製することが可能である。蛍光原位置ハイブリッド形成法(FISH)は、Heinz-Ulrichら(In: Meyers前出965-968ページ)に記載されているように他の物理的染色体マッピング技術及び遺伝地図データと相関し得る。

20

【0062】

別の実施例では、タンパク質の過剰発現または過少発現によって特徴付けられる疾患の診断と予後のために、タンパク質を特異結合する抗原結合領域部位を含む抗体またはFabを使用することができる。タンパク質発現を測定するための様々なプロトコル、例えばELISA、RIA、FACS等が本技術分野において知られており、発現の変異レベル或いは異常レベルを診断する基準を提供する。化合物形成の条件下で、健康な被験者、好ましくはヒトから採取したサンプルをタンパク質への抗体と比較してタンパク質発現の標準値を確立する。様々な方法、好ましくは測光法により化合物形成の量を定量することは可能である。疾患サンプルで発現したタンパク質の量を基準値と比較する。標準値と被験者との偏差によって、疾患を診断またはモニターするパラメータが確立される。または、タンパク質を特異結合することができる中和抗体が試験化合物と競合する、競合的薬剤スクリーニングアッセイを使用することが可能である。1個若しくは数個の抗原決定因子をタンパク質と共有する任意のペプチドの存在を検出するために、抗体を使用することが可能である。或る実施態様において、細胞周期疾患の治療または治療上の処置をモニターするために抗体を使用することができる。

30

【0063】

別の実施態様では、mRNAとタンパク質を発現する目的、または逆にmRNAの転写あるいは翻訳を阻害する目的で、cDNAまたはその相補体を治療上使用することができる。レトロウイルス、アデノウイルス、ヘルペスウイルスまたはワクシニアウイルス由来のまたは細菌性プラスミド等由来の要素を使用して発現ベクターを作製することができる。これらのベクターを使用して、ヌクレオチド配列を特定の標的器官、組織または細胞集団へ送達することができる。当業者によく知られている方法を用いて、核酸配列またはその相補体を発現するためにベクターを作製することが可能である。(Maulikら(1997) Molecular Biotechnology, Therapeutic Applications and Strategies, Wiley-Liss, New York NY等を参照)。また体細胞または幹細胞の遺伝子治療のためにcDNA

40

50

あるいはその相補体を使用することができる。ベクターを *in vivo*、*in vitro*、*ex vivo* に導入することが可能である。*ex vivo* 治療の場合、ベクターを患者から採取した幹細胞内に導入し、得られた遺伝形質転換細胞をクローニング増殖して同一患者に自家移植で戻ることができる。トランスフェクション、リボソーム注入またはポリカチオンアミノポリマーによる cDNA の輸送は、当分野でよく知られている方法を用いて実行することができる (Goldman ら (1997) *Nature Biotechnology* 15: 462 - 466 等を参照)。さらに、不活性遺伝子配列を cDNA のコードする領域または他の標的領域に挿入する相同組換え方法を使用して内在性遺伝子発現を不活性化することができる (Thomas ら (1987) *Cell* 51: 503 - 512 等を参照)。

10

【0064】

欠損したタンパク質を発現するため、または機能しないタンパク質と取り替えるために、細胞または組織へと cDNA を含むベクターを形質転換することは可能である。同様に、cDNA の相補体を発現するよう作製されたベクターをタンパク質発現を下方調節するため細胞に形質転換することができる。相補的またはアンチセンス配列は、好適には A T G から約 10 と + 10 の間の位置で転写開始部位由来のオリゴヌクレオチドから成ることも可能である。同様に、三重らせん塩基対の形成方法を用いて阻害が可能となる。三重らせん塩基対形成は、ポリメラーゼ、転写因子、エンハンサーまたは調節分子の結合のために十分に開くような二重らせんの能力を阻害するので、三重らせん塩基対形成は有用である。三重らせん DNA を用いる最近の治療の進歩については文献に記載がある (Gee ら In: Huber 及び Carr (1994) *Molecular and Immunologic Approaches*, Futura Publishing, Mt. Kisco NY、163 - 177 ページ当を参照)。

20

【0065】

リボザイムは酵素性 RNA 分子であり、本発明の cDNA から成る mRNA の切断を触媒するため、また特定の mRNA のレベルを低下するためにリボザイムを用いることもできる。(Rossi (1994) *Current Biology* 4: 469 - 471 等を参照)。リボザイムは、mRNA を特異的切断部位で切断することが可能である。またはリボザイムは、標的 mRNA と共に相補的塩基対を形成する側面に位置する領域により影響を受ける場所で mRNA を切断することが可能である。リボザイムの作製と生産は当分野で公知であり、Meyers (前出) に記載されている。

30

【0066】

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。限定するものではないが可能な修飾には、分子の 5' 末端、3' 末端、あるいはその両方においてフランキング配列を追加したり、分子の主鎖内においてホスホジエステラーゼ結合ではなくホスホジエステルまたは 2' - O - メチルを使用したりすることが含まれる。または内在性エンドヌクレアーゼによって容易には認識されないアデニン、シチジン、グアニン、チミン、及びウリジンにアセチル -、メチル -、チオ - 及び同様の修飾をしたものの他、非従来型塩基、例えばイノシン、クエオシン (queosine)、ワイプトシン (wybutosine) 等を含めることができる。

40

【0067】

さらに細胞周期疾患を治療する患者に、タンパク質を特異結合するアンタゴニストまたは抗体を投与することも可能である。タンパク質の活性を阻害するために直接、あるいはタンパク質を発現するために薬剤を細胞か組織に送達するために間接的にアンタゴニスト、抗体または断片を使用することができる。薬剤は、以下の群から選択した細胞毒でありうる。限定するものではないが、その群にはアプリン、リシン、ドキシソルピシン、ダウノルピシン、タキソール、臭化エチジウム、マイトマイシン、etoposide、tenoposide、ピンクリスチン、ピンブラスチン、コルヒチン、ジヒドロキシアントラシンジオン、アクチノマイシン D、ジフテリア毒素、シュードモナス外毒素 A、そして、40 放射性同位元素そして、グルココルチコイドが含まれる。

50

【0068】

タンパク質の抗体も、当分野で一般的な方法を用いて製造することが可能である。このような抗体には、ポリクローナル抗体、モノクローナル抗体、キメラ抗体、一本鎖、Fabフラグメント、及びFab発現ライブラリによって作られたフラグメントが含まれる。但し、これらに限定されるものではない。中和抗体（即ち二量体の形成を阻害する抗体）は通常、治療用に好適である。タンパク質に対するモノクローナル抗体は、抗体分子を産生する任意の技術を用いて、培地内の連続した細胞株によって作製し得る。限定するものではないがこのような技術には、ハイブリドーマ技術、ヒトB細胞ハイブリドーマ技術及びEBV-ハイブリドーマ技術がある更に、キメラ抗体作製のために発達した技術を使用することができる。（Pound（1998）Immunochemical Protocols, Methods Mol Biol Vol. 80等を参照）。別法では、一本鎖抗体の産生のための記載された技術を使用できる。タンパク質に対する特異的な結合部位を含むFabも生産することができる。種々のイムノアッセイを用いて所望の特異性を有する抗体を同定することができる。隔離された特異性を有するポリクローナル抗体またはモノクローナル抗体の何れかを用いる競合的な結合、または免疫放射線アッセイのための数々のプロトコルが、当分野では周知である。

10

【0069】

更にタンパク質の発現低下、寿命または活性に関連した疾患の治療または予防のために、タンパク質のアゴニストを患者に投与することも可能である。

【0070】

本発明の更なる実施形態では、すでに記載した任意の治療の適用のために薬剤として許容できる担体と併用して薬剤または滅菌された組成剤の投与に関連する。このような医薬品成分は、タンパク質または抗体、擬態、アゴニスト、アンタゴニスト、またはタンパク質のインヒビターから構成し得る。組成物は単独で、あるいは安定化剤など少なくとも他の一つの薬剤と併用して投与することが可能であり、限定するものではないが、生理食塩水、緩衝食塩水、ブドウ糖、水等の滅菌された生体適合性のある担体で投与される。組成物は単独で、あるいは他の薬品、薬剤、ホルモンと併用して患者に投与することが可能である。

20

【0071】

本発明に用いられる医薬品成分は、任意の数の経路によって投与することができ、限定するものではないが経路には、経口、静脈内、筋肉内、動脈内、骨髄内、クモ膜下腔内、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。活性成分の他に、これらの医薬品成分は、薬剤として使用できる医薬品へと活性成分の処理を促進する賦形剤と補助剤を含む、薬剤として許容できる担体を含有することが可能である。製剤法と投与方法に関する技術上の詳細は、Remington's Pharmaceutical Sciences（Maack Publishing, Easton PA）の最新版に見出すことができる。

30

【0072】

任意の化合物に対して、細胞培養アッセイ或いは、動物モデル、例えばマウス、ラット、ウサギ、イヌ、ブタ等において、先ず治療の有効投与量を推定することができる。動物モデルはまた、濃度範囲及び投与経路を決定するためにも用い得る。このような情報を用いて、次にヒトに対する有益な投与量及び投与経路を決定することができる。

40

【0073】

医学的に効果的な薬用量は、症状や容態を回復させる活性処方成分の量に関連する。薬用有効度及び毒性は、細胞培養または動物実験における標準的な薬剤手法によって、例えばED₅₀（集団の50%の医薬的有効量）またはLD₅₀（集団の50%の致死量）統計を測定、対比するなどして決定することができる。上記の医薬品組成物はいずれも、限定するわけではないが、イヌ、ネコ、ウシ、ウマ、ウサギ、サルそして最も好適なヒト等の哺乳動物を含めて治療が必要な全ての対象に適用できる。

【0074】

50

(実施例)

本発明が、説明した特定の装置、機械、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。特定の実施例が記述されていても、同等な実施例が本発明を実施するために使用されうる。記載した実施例は本発明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。

【0075】

1 cDNAライブラリ作製

肺の右上葉の悪性腫瘍診断の後に、68歳の白人男性の肺の部分切除中に採取した癌性肺組織からLUNGTUT09cDNAライブラリを構築した。肺の右上葉の病理は、気管と周囲の柔組織にも影響する浸潤塊を形成する侵襲性グレード3扁平上皮癌を示した。患者の病歴には、以前に診断されていた、合併症を併発しないII型糖尿病、甲状腺障害、鬱病、高脂血症、食道の潰瘍とアテローム性動脈硬化症が含まれる。家族歴には、母親と父親のアルコールの使用、兄弟と祖父母にアテローム性動脈硬化症、母親に悪性脳腫瘍があった。

10

【0076】

POLYTRONホモジナイザー(Brinkmann Instruments, Westbury NY)を使用して、凍結組織をホモジナイズして、TRIZOL試薬(1g tissue/10ml, Life Technologies)に溶解する。氷上で短いインキュベーションの後に、クロロホルムを加え(1:5 v/v)、溶解物を遠心分離機にかけた。クロロホルム上層を新しい試験管に移し、イソプロパノールで抽出したRNAをDEPC処理水で再懸濁し、37°Cで25分間DNアーゼ処理した。RNAをpH 4.7の酸性フェノール-クロロホルムで再抽出し、0.3Mの酢酸ナトリウムと2.5倍量のエタノールを用いて沈殿した。mRNAをOLIGOTEXキット(Qiagen, Chatsworth CA)で単離して、cDNAライブラリの作製のために用いた。

20

【0077】

mRNAをSUPERSCRIPTプラスミドシステム(Life Technologies)で推奨されたプロトコルに従って処理した。cDNAをSEPHAROSE CL4Bカラム(Amersham Pharmacia Biotech (APB), Piscataway NJ)上で分画して、そのうち400bpを超えるcDNAはpINCYプラスミド(Incycyte Genomics, Palo Alto CA)に連結反応させた。プラスミドは、次にDH5a'大腸菌細胞(Life Technologies)に形質転換させた。

30

【0078】

2 cDNAクローンの単離とシーケンシング

プラスミドDNAを細胞から放出して、REAL PREP 96 プラスミドキット(Qiagen)を用いて精製した。推奨されているプロトコルを下記の変更を除いて使用する。1)細菌をカルベニシリン25 mg/lと0.4%のグリセロールで、殺菌したTERRIFIC BROTH(BD Biosciences, San Jose CA) 1 mlで培養した。2)ウェルを接種後、培養を19時間インキュベートし、次いで0.3 mlの溶解バッファで溶解した。そして3)イソプロパノール沈殿後、DNAペレットを0.1 mlの蒸留水で再懸濁した。プロトコルの最終ステップ後に、サンプルを96-ウェルブロックに移して、4°Cで保管した。

40

【0079】

cDNAをDNA ENGINE サーマルサイクラー(MJ Research, Watertown MA)と併用してMICROLAB 2200システム(Hamilton, Reno NV)を使用して調製した。cDNAをABI PRISM 377シーケンシングシステム(ABI)を用いてSangerとCoulson(1975; J Mol Biol 94:441f)の方法で配列化した。ほとんどの配列は、0.25x~1.0xの溶液量で標準ABIプロトコルとキット(ABI)を用いて配列化した。別法

50

では、幾つかの配列を A P B の溶液と色素を用いて配列化した。

【 0 0 8 0 】

3 配列の選択、構築、性質決定

同時発現分析に用いられた配列は、E S T 配列、5' 及び 3' ロングリードシーケンス、全長コード配列から構築した。選択した構築配列を少なくとも 3 つの c D N A ライブラリで発現した。

【 0 0 8 1 】

構築プロセスは下記の通りである。E S T 配列クロマトグラムを処理して、検証した。P H R E D (E w i n g ら (1 9 9 8) G e n o m e R e s 8 : 1 7 5 - 1 8 5、E w i n g 及び G r e e n (1 9 9 8) G e n o m e R e s 8 : 1 8 6 - 1 9 4) を用いて、品質スコアを得た。そして編集した配列をリレーショナル・データベース管理システム (R D B M S) にロードする。配列を積スコア 5 0 で B L A S T によりクラスタ化した。2 以上の配列のすべてのクラスタは、1 つの転写遺伝子を代表するピンを作成した。

10

【 0 0 8 2 】

D N A 断片を構築するために一般に入手可能なプログラムである P h r a p の修飾を利用して、各ピン内の部分配列の構築を実施した (G r e e n , P . U n i v e r s i t y o f W a s h i n g t o n , S e a t t l e W A) 。すべてのコンセンサス配列間の局所ペアワイズアラインメントで、8 2 % の同一性を示したピンを組み合わせた。

N C B I の G B p r i と G e n P e p t 等の公共のデータベースに対し、各ピンでコンセンサス配列をスクリーニングすることによりピンに注釈を付けた。注釈を付ける処理には、G e n B a n k の G B p r i データベースに対して F A S T n スクリーニングすることが関係する。7 5 % 以上の一致率を有するヒット、およびアラインメント長さが 1 0 0 塩基対以上であるアラインメントは、相同体ヒットとして記録した。G e n P e p t に対する F A S T x により残りの注釈を付けていない配列をスクリーニングした。1 0⁻⁸ 以下の E 値を有するこれらのヒットを相同体ヒットとして記録した。

20

【 0 0 8 3 】

次いで、迅速なアミノ酸と核酸配列の比較及びデータ・ベース検索用プログラムである B L A S T n と C r o s s - M a t c h (前出の G r e e n) を使用して配列を連続して再びクラスタ化した。スコアが 1 5 0 超である配列とコンセンサス配列間のすべての B L A S T アラインメントに対して c r o s s - m a t c h を利用して、再びアラインメントする。コンセンサス配列が少なくとも 8 2 % の同一性を示す局所アラインメントの中で最高の S m i t h - W a t e r m a n スコアを出したピンに、配列を加えた (S m i t h ら (1 9 9 2) P r o t e i n E n g i n e e r i n g 5 : 3 5 - 5 1) 。適合しない配列を新しいピンに移して、構築プロセスを繰り返した。

30

【 0 0 8 4 】

4 既知の細胞周期遺伝子の説明

細胞周期に関連する疾患プロセスに関与しているとして知られている遺伝子を選択して、c D N A を同定した。既知の遺伝子と、機能の簡単な説明を下記に記す。

【 外 1 】

40

遺伝子ID	名前	説明	
995529	CDC2	細胞分裂周期タンパク質2 (あるいはサイクリンB1)であるCDC2は、有糸分裂キナーゼである。また有糸分裂を誘発する。CDC2はS期前にクロマチンを結合し、DNA複製中にはずれる(Krude <u>ら</u> (1996) J Cell Sci 109:309-318、De Souza <u>ら</u> (2000) Exp Cell Res 257:11-21)。	
336106	CDC7	細胞分裂周期タンパク質7、CDC7は酵母からヒトまでの真核生物で保存されたキナーゼである。それはDNA複製の開始とS期へ移行にとって不可欠である(Donaldson <u>ら</u> (1998) Genes Dev 12:491-501、Jiang <u>ら</u> (1999) Embo J 18: 5703-5713、Masai <u>ら</u> (1999) Front Biosci 4: D834-840)。	10
256671	CDC23	細胞分裂周期タンパク質23であるCDC23は、サイクリンBユビキチン抱合体の形成を触媒すること及び分解のためにサイクリンBを標的化することにより、有糸分裂を調節する後期促進複合体の成分である(Prinz (1998) Curr Biol 8:750-760、Zhao <u>ら</u> (1998) Genomics 53:184-90、Hershko (1999) Philos Trans R Soc Lond B Biol Sci 354:1571-1576)。	
286623	サイクリンB	サイクリンBは、サイクリン依存性キナーゼ(Cdk)のサブユニットである。後期促進複合体によるサイクリンBの分解は、キナーゼの不活性化と有糸分裂から出るために不可欠である。CDKは細胞周期の進行と変更の制御因子であり、CDK活性の変更および調節不全は腫瘍形成の特徴である。CDKインヒビターとモジュレーターは細胞周期を変え、またアポトーシスと腫瘍の退行を誘発する(Hajdуч <u>ら</u> (1999) Adv Exp Med Biol 457:341-53、前出のHershko及びSausville (1999) Pharmacol Ther 82:285-92)。	20

【外 2】

遺伝子ID	名前	説明	
392739	hBub1	有糸分裂チェックポイントキナーゼであるhBub1は、有糸分裂細胞で染色体の紡錘体への付着をモニターする、また有糸分裂から出ることと染色体分離を制御する動原体タンパク質である。有糸分裂チェックポイントは、紡錘体上で染色体がアラインメントされるまで後期を遅延することにより適切な染色体分離を確実にする。紡錘体の損傷後には、細胞は有糸分裂を出て、アポトーシスを受ける。hBub1は、紡錘体の損傷へのチェックポイント反応のために必要である。hBub1の突然変異は、異数性につながり、腫瘍形成に寄与する可能性がある紡錘体の損傷にもかかわらず、細胞がアポトーシスを逃れ、細胞周期進行を継続することを可能にする有糸分裂チェックポイントを破壊する(Taylor及びMcKeon (1997) Cell 89:727-735, Cahill (1998) Nature 392:300-303, Ouyang <u>ら</u> (1998) Cell Growth Differ 9:877-885, Imai <u>ら</u> (1999) Jpn J Cancer Res 90:837-840, Seeley <u>ら</u> (1999) Biochem Biophys Res Commun 257:589-595, Myrie <u>ら</u> (2000) Cancer Lett 152:193-99)。	10
337334	hKSP	キネシン様紡錘体タンパク質(HsEg5)であるhKSPは、有糸分裂後の中心体移動と関連する前期、前中期の中心体分離中に中心体の微小管と共に見出される紡錘体関連タンパク質である(Whitehead <u>ら</u> (1996) Cell Motil Cytoskeleton 35:298-308)。	20
201204	hp55cdc	hp55cdcは、紡錘体チェックポイントタンパク質Mad2とcyclosome/後期促進化合物の結合を仲介する紡錘体と動原体微小管関連タンパク質であり、細胞分裂に不可欠である。p55cdcの過剰発現はアポトーシスを誘発する。hp55cdcは有糸分裂紡錘体タンパク質キナーゼAikとも会合する(Weinstein <u>ら</u> (1994) Mol Cell Biol 14:3350-3363, Kao <u>ら</u> (1996) Oncogene 13:1221-1229, Kallio <u>ら</u> (1998) J Cell Biol 141: 1393-1406, Kramer <u>ら</u> (1998) Curr Biol 8:1207-1210; Farruggio <u>ら</u> (1999) Proc Natl Acad Sci 96:7306-7311, Saffery <u>ら</u> (2000) Hum Mol Genet 9:175-85)。	30
331025	MCAK	有糸分裂のセントロメア結合キネシンであるMCAKは、前期にセントロメアに動員される微小管モータータンパク質であり、後期染色体分離に関与する(Kim <u>ら</u> (1997) Biochim Biophys Acta 1359:181-186, Maney <u>ら</u> . (2000) Int Rev Cytol 194:67-131, Maney <u>ら</u> (1998) J Cell Biol 142:787-801, Wordeman <u>ら</u> (1999) Cell Biol Int 23:275-86, 前出のSaffery)。	

【外3】

遺伝子ID	名前	説明	
26662	mitosin	Mitosin (CENP-F 動原体タンパク質) は、M期中にセントロメアと紡錘体極と関連する核タンパク質である。N-末端が切り取られたmitosinの過剰発現により細胞周期進行を妨げる。Mitosinはリンパ節転移陰性乳癌の臨床結果と相関する(Clark <u>ら</u> , (1997) <i>Cancer Res</i> 57:5505-08, Zhu (1999) <i>Mol Cell Biol</i> 19:1016-1024, Zhu <u>ら</u> (1997) <i>J Cell Biochem</i> 66:441-449)。	
412661	mki67a	mki67a (MIB-1)は、決定的な細胞増殖マーカーである。病理学では、ヒトの腫瘍の細胞の増殖分画を測定するために広く使用されている(Schluter <u>ら</u> (1993) <i>J Cell Biol</i> 123:513-522, Duchrow <u>ら</u> (1995) <i>Arch Immunol Ther Exp</i> 43:117-121, Dalquen <u>ら</u> (1997) <i>Acta Cytol</i> 41:229-237, Scholzen及びGerdes (2000) <i>J Cell Physiol</i> 182:311-322)。	10
319885	MKLP-1	有糸分裂キネシン様タンパク質1であるMKLP1は、有糸分裂進行のために必要な紡錘体関連タンパク質である(Nislow <u>ら</u> (1992) <i>Nature</i> 359:543-7, Sharp <u>ら</u> (1997) <i>J Cell Biol</i> 138:833-843, Kobayashi <u>ら</u> (1998) <i>J Cell Biol</i> 143:1961-70)。	
977509	myb	B-mybは細胞周期調節転写因子であるmybファミリーのメンバーであり、G1とS期に発現する。b-mybの活性はサイクリンA/cdk2依存リン酸化により刺激される(Robinson <u>ら</u> (1996) <i>Oncogene</i> 12:1855-64, Saville及びWatson (1998) <i>Adv Cancer Res</i> 72:109-40, Saville及びWatson (1998) <i>Oncogene</i> 17:2679-2689, Horstmann <u>ら</u> (2000) <i>Oncogene</i> 19:298-306)。	20
336560	NLK1	NIMA様タンパク質キナーゼである1NLK1は、アスペルギルスのNIMA細胞周期調節タンパク質キナーゼに類似のヒト有糸分裂キナーゼであり、有糸分裂の誘発と進行に不可欠である(Lu及びHunter (1995) <i>Cell</i> 81:413-424, Lu及びHunter (1995) <i>Prog Cell Cycle Res</i> 1:187-205, Shen <u>ら</u> (1997) <i>Proc Natl Acad Sci</i> 94:13618-13623)。	30
347876	P1-CDC21	P1-CDC21は、ミニクロモゾーム維持タンパク質のファミリーのメンバーであり、DNAの複製に不可欠である(Hu <u>ら</u> , (1993) <i>Nucleic Acids Res</i> 21:5289-5293, Ishimi <u>ら</u> (1996) <i>J Biol Chem</i> 271:24115-24122)。	
411205	PRC1	タンパク質調節細胞質分裂1であるPRC1は、細胞質分裂に必要なヒト有糸分裂紡錘体関連CDK基質タンパク質である(Jiang <u>ら</u> (1998) <i>Mol Cell</i> 2:877-885)。	

【 外 4 】

遺伝子ID	名前	説明	
348211	Aik2	プロテインキナーゼAik2/Aurora2は、有糸分裂紡錘体極に局在しており、また染色体分離の調節とゲノム安定性の維持に関与しており、p55cdc/cdc20と関連する(Kimura <u>ら</u> (1999) J Biol Chem 274:7334-40、Kimura <u>ら</u> (1998) Cytogenet Cell Genet 82:147-52、前出のFarruggio)。	
251651	サバイピン	サバイピンは、細胞周期のG2/M期に発現するアポトーシス阻害剤である。有糸分裂の開始において紡錘体の微小管と会合する。アポトーシスを阻害して、癌細胞の存続を可能にする(Li <u>ら</u> (1998) Nature 396:580-584、Verdecia <u>ら</u> (2000) Nat Struct Biol 7:602-608)。	10
232888	topo II	トポイソメラーゼII型は、DNAの複製中に染色体の凝縮と分離に必要である。発現は細胞周期に依存しており、タンパク質レベルと触媒作用の両者はG2/M期でピークに達する。有糸分裂の誘発と進行の調節チェックポイントの一部として、アポトーシスを調節する。トポイソメラーゼ毒は発癌性の染色体変異を誘導する(Holm <u>ら</u> (1989) Mol Cell Biol 9:159-168、Kaufmann (1998) Proc Soc Exp Biol Med 217:327-334、Sumner (1995) Exp Cell Res 217:440-447、Anderson及びRoberge (1996) Cell Growth Differ 7:83-90、Larsen <u>ら</u> (1996) Prog Cell Cycle Res 2:229-239、Cimini <u>ら</u> (1997) Cytogenet Cell Genet 76:61-67)。	20
235191	UbcH10	サイクリン選択性ユビキチンキャリアータンパク質(UbcH10/E2-C)は、有糸分裂サイクリンのユビキチン仲介蛋白質分解に触媒作用を及ぼす。また細胞が有糸分裂を完了して、次の細胞周期の後期を誘発するために必要である。突然変異UbcH10はサイクリンの破壊を阻害し、M期で細胞を停止する。そして後期の開始を阻害する(Townesley <u>ら</u> (1997) Proc Natl Acad Sci 94:2362-2367、Bastians <u>ら</u> (1999) Mol Biol Cell 10:3927-3941)。	

【0085】

30

5 既知の細胞周期遺伝子の同時発現分析

L I F E S E Q G O L D データベース (D e c 9 9 , I n c y t e G e n o m i c s) を使用して、既知の細胞周期遺伝子と強い結合を示す10のcDNAを同定した。始めに、0.00001未満のカットオフp値を使用して確率値により、関連性の度合いを測定した。続いて、確率テストを通過する遺伝子が既知の細胞周期遺伝子と強い関連性を確実に有するようにするためにアノテーションと文献検索を実施した。プロセスを繰り返して、37,071の遺伝子からの最初の選択を最終的に本明細書で請求している10のcDNAに減らした。下の表1のエントリは、二つの遺伝子の同時発現のためのp値の対数の負をとった数($-\log p$)である。L I F E S E Q G O L D I D 番号によりcDNAを同定して、既知の遺伝子を上記のように略語で同定した。そして行1でも用いられている番号を列1にも割り当てた。各既知の遺伝子間で最高のp値を太字でマークした。少なくとも1つの既知の遺伝子と各cDNA間の最高のp値を「発明」のセクションで要約した。

40

【0086】

【表1】

名/番号	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
	1																		
1 CDC2	NA																		
2 CDC23	13	NA																	
3 CDC7	3.4	5.5	NA																
4 サイクリンB	12	6.6	10	NA															
5 hBub1	7.8	7.7	0	5.2	NA														
6 hKSP	5.8	6	5.2	10	4.9	NA													
7 hp55cdc	5.7	4.8	5.8	12	5.5	7.2	NA												
8 MCAK	4.8	6.9	9.2	14	6.1	8	11	NA											
9 mitosin	9.8	4.6	6.6	14	12	7	13	12	NA										
10 mki67a	12	5.5	8.3	6.7	6.5	5.8	7.8	11	13	NA									
11 MKLP-1	6.9	5.1	4.5	3.8	6.9	5.3	8.8	3.9	7.1	7.2	NA								
12 myb	0	5.5	7.9	13	5.9	5.7	19	18	13	20	6.8	NA							
13 NLK1	0	4.3	0	0	10	3.5	4.2	4.8	8.3	3.9	3.9	0	NA						
14 P1-CDC21	11	7.4	7	10	6.4	5.6	14	9.2	12	21	5.5	10	5.6	NA					
15 PRC1	7.2	5.4	5.9	15	9.7	12	16	12	13	13	4.6	8.5	7.4	15	NA				
16 prkAik2	6.6	8.4	3.2	11	5	6.8	9.3	6.7	7.6	11	5	8.1	3.8	7.2	10	NA			
17 サバイピン	11	6.5	9.2	11	9.3	9.2	18	11	18	9	10	11	6.1	11	7.9	9.8	NA		
18 topo II	23	11	13	17	11	15	19	14	24	15	6.1	18	8.2	19	18	11	23	NA	
19 UbcH10	6.1	5.8	4	12	7.5	12	15	8.7	12	12	5.6	16	6.2	9.4	8	15	16	11	NA
40371	7.9	3.9	5.8	4.6	9.4	6.6	5.5	7.6	7.6	8.8	7.5	7	9.1	8.5	8.7	6.4	5.8	16	10
200394	7.1	5.2	4.9	7.6	6.5	7.5	8.9	6.8	4.5	7.3	9	4.3	4.2	9.5	12	6.6	5.2	11	7.1
201989	5.9	12	4.5	10	4.5	5.4	8.3	9.5	6.7	9.6	3.2	9.7	0	11	8.5	11	8	11	11
211475	9.2	6.6	4.8	7.9	4.5	5.9	8.2	6	5.5	6.7	4.3	4.9	0	9.8	10	7.6	5.4	10	6.2
225657	4.7	5	5	13	4.5	4.5	7.9	9.8	11	8.3	3.7	7.2	3.7	6.4	5.5	11	8.4	6.7	13
350770	6.5	6.5	9.9	7.2	5.8	5.9	12	11	9.1	12	0	11	3.7	15	16	8	6.5	14	11
407614	6.6	0	5.5	9.5	0	3.6	7.9	7.6	4.9	4.8	3.8	6.5	0	3.5	5.9	4.6	8.2	9.4	5.6
475113	9	10	4.2	6.7	7.6	8.6	9.1	5.1	9.3	9.4	5.7	10	4.5	9	10	8.4	9	13	12
898622	0	0	0	5.2	5.6	7.9	3	0	4.9	0	0	0	3.3	5.2	4.2	4	6.6	9	3.2
978267	0	3.4	10	7.8	3.8	5	17	9.5	8	4.1	4.2	0	4.2	7.8	8.2	5.6	9.4	15	5.6

10

20

30

40

50

【 0 0 8 7 】

6 cDNAクローンの相同性検索と推定タンパク質

配列表のcDNAまたは推定アミノ酸配列を用いて、GenBank、SwissProt、BLOCKSなどのデータベースに問合せした。BLASTかBLAST2（前出のAltschulら；前出のAltschul）を用いて前もって同定し、またアノテーションが付けられた配列あるいはドメインを含むこれらのデータベースを検索して、アラインメントを作製し、どの配列が厳密に一致するか、または相同的かを決定した。アラインメントを原核（細菌）または真核（動物、カビあるいは植物）生物のシークエンシングした。別法として、SmithとSmith（1992, Protein Engineering 5: 35-51）で説明されているようなアルゴリズムを用いて、一次配列パターンと二次構造ギャップペナルティで処理することが可能であった。この出願で開示されているすべての配列は、少なくとも49ヌクレオチドの長さを有し、不要な塩基は1.2%に過ぎない（A、C、GまたはTよりもNが記録されている）。

【 0 0 8 8 】

前出のKarlinで詳説されているように、問合せ配列とデータベース配列の間のBLAST一致を統計的に評価して、ヌクレオチドに $10^{-2.5}$ 、ペプチドに $10^{-1.4}$ の閾値を満たす時のみ報告した。下記のように計算して積スコアによって、相同性も評価した。BLASTのヌクレオチドあるいはアミノ酸の一致率[問い合わせ配列と参照配列間]に%最大限BLASTスコアを乗じ[問い合わせ配列と参照配列の長さに基づく]、次いで100で除する。実験室で用いるハイブリダイゼーション法と比較して、厳密な一致の電子ストリンジェンシーを約40の下限（不要な塩基のために1-2%のエラーがある）か

ら70の100%一致まで設定した。

【0089】

無料で入手可能な配列比較アルゴリズムであるBLASTソフトウェア一式(NCBI, Bethesda MD, <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>)には、核酸分子をアラインメントする「blastn」、核酸分子またはアミノ酸分子を対で直接比較するために用いるBLAST 2など多種の配列分析プログラムが含まれる。BLASTプログラムは、一般的には、ギャップ及びデフォルト設定に設定された他のパラメータと共に用いる。例えば: Matrix: BLOSUM62; Reward for match: 1; Penalty for mismatch: -2; オープンギャップ5 及びExtension Gap: 2 ペナルティ; ギャップ x drop-off: 50; Expect: 10; Word Size: 11; 及びFilter: on同一性または類似性は配列の全長またはより小さいその部分の長さにより測定する。Brenner ら (1998; Proc Natl Acad Sci 95: 6073 - 6078, 文献は、引用することをもって本明細書の一部とする)は配列同一性によって構造的相同性を同定する能力に関してBLASTを分析した。少なくとも150残基の配列アラインメントには信頼できる閾値は30%の同一性であり、少なくとも70残基のアラインメントに関しては40%であることがわかった。

10

【0090】

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

20

【0091】

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

30

【0092】

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

40

【0093】

構築に続いて、鋳型をBLAST、モチーフその他の機能的分析にかけて、1997年3

50

月6日に提出したUSSN 08/812, 290、USSN 08/811, 758、1997年10月9日に提出したUSSN 08/947, 845、1998年3月4日に提出したUSSN 09/034, 807等に記載されているタンパク質階層に分類した。3つの全フォワードリーディングフレームで各鋳型を翻訳することにより、またHMMER ソフトウェアパッケージ (Washington University School of Medicine, St. Louis MO; <http://pfam.wustl.edu/>) を用いて隠れマルコフモデル (HMM) を基にした保存されたタンパク質ファミリードメインのPFAMデータベースに対する各翻訳を検索することにより鋳型を分析した。

【0094】

MACDNASIS PROソフトウェア (Hitachi Software Engineering) とLASERGENEソフトウェア (DNASTAR) を用いてcDNAを更に分析した。そして公共のデータベース、例えばGenBankのげっ歯類、哺乳動物、脊椎動物、原核生物、真核生物のデータベースと、SwissProt、BLOCKS、PRINTS、PFAMそしてPrositateに問い合わせた。

【0095】

7 染色体マッピング

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

【0096】

8 ハイブリダイゼーション技術と分析

基質上でのcDNAの固定

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

【0097】

2番目の方法では、インサートに隣接するベクター配列に相補的なプライマーを用いて、cDNAを30サイクルのPCRで細菌ベクターから増幅する。PCR増幅で1~2ngの初期量の核酸から5µgより大きい最終量まで増大する。SEPHACRYL-400ビーズ (APB) を用いて長さが400 bpから約5000 bpまで増幅した核酸を精製する。手であるいはドット/スロットプロット法マニホールドと吸気装置を用いて精製核酸をナイロン膜に置く。そして上述した変性、中和、UV照射により固定する。USPN 5, 807, 522に記載されている方法を用いて、精製した核酸を、ロボットで配置して、ポリマーコートされたスライドガラス上に固定する。顕微鏡スライドガラス (Corning, Acton MA) を良く洗って0.1%のSDS中で超音波をかけ、4%フッ化水素酸 (VWR Scientific Products, West Chester PA) 中でエッチングし、95%エタノール中で0.05%アミノプロピルシ

10

20

30

40

50

ラン (Sigma Aldrich) を用いてコーティングする、そして 110 のオープンで硬化させることにより、ポリマーコートされたスライドガラスを準備する。スライドガラスを処理中と後で蒸留水中で広範囲にわたって洗浄する。核酸をスライドガラス上に置き、次に STRATALINKER UV 架橋剤 (Stratagene) を用いてアレイを UV 照射に暴露することにより固定する。アレイを室温において 0.2% SDS で洗浄し、蒸留水で 3 度洗浄する。リン酸緩衝生理食塩水 (PBS) (Tropix, Inc., Bedford MA) 中の 0.2% カゼイン中において 60 で 30 分間アレイをインキュベートした後、0.2% SDS 及び蒸留水で洗浄することにより、非特異結合部位をブロックする。

【0098】

膜ハイブリダイゼーションのプロープ調製

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

【0099】

ポリマーコートされたスライドハイブリダイゼーションのプロープ調製

サンプルから分離した mRNA に由来するハイブリダイゼーションのプロープを使用して、アレイベースのハイブリダイゼーションで配列表の cDNA をスクリーニングする。9 μ l TE バッファーで濃度 200 ng に mRNA を希釈し、5 μ l 5x バッファー、1 μ l 0.1 M DTT、3 μ l Cy3 または Cy5 標識化混合液、1 μ l RNアーゼ阻害因子、1 μ l 逆転写酵素、5 μ l 1x 酵母調節 mRNA を加え、GEMbright キット (Incyte Genomics) を用いてプロープを調製する。非コード酵母ゲノム DNA から *in vitro* 転写により酵母調節 mRNA を合成する (W. Lei, 未発表)。定量用対照として、サンプル mRNA に 0.002 ng、0.02 ng、0.2 ng、2 ng での対照 mRNA の 1 つの設定をそれぞれ 1:100, 000、1:10,000、1:1000、1:100 (w/w) の比で逆転写反応混合液に希釈する。mRNA 示差発現パターンを調べるために、対照 mRNA の 2 番目の設定を 1:3, 3:1, 1:10, 10:1, 1:25, and 25:1 (w/w) の比で逆転写反応混合液に希釈する。反応混合液を混合して、37 で 2 時間インキュベートする。次に反応混合液を 20 分間、85 でインキュベートする。そして 2 つの連続する CHROMA SPIN+TE 30 カラム (Clontech, Palo Alto CA) を利用してプロープを精製する。精製したプロープは、プロープを DEPC 処理水の 90 μ l に希釈させ、2 μ l 1mg/ml グリコーゲン、60 μ l 5 M 酢酸ナトリウム、300 μ l 100% エタノールを加えて、エタノールで析出させた。プロープを 20 分間、20,800 x g で遠心分離する。そしてペレットを 12 μ l の再懸濁バッファーで再懸濁し、5 分間 65 に加熱してから、十分に混合する。プロープを加熱して、以前のように混合して、氷上に保管する。以下に詳述するように、プロー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。

【0100】

膜系ハイブリダイゼーション

1% Sarkosyl と 1x 高リン酸バッファ (0.5 M NaCl, 0.1 M Na₂HPO₄, 5 mM EDTA, pH 7) を含むハイブリダイゼーション溶液で、

10

20

30

40

50

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

【0101】

ポリマーコートされたスライドハイブリダイゼーション

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

【0102】

ハイブリダイゼーション反応を絶対的または示差ハイブリダイゼーションフォーマットで実施する。絶対ハイブリダイゼーションフォーマットでは、1つのサンプルのプローブをアレイエレメントにハイブリダイズし、ハイブリダイゼーション複合体形成後のシグナルを検出する。シグナル強度は、サンプル中のプローブmRNAレベルと相関する。示差ハイブリダイゼーションフォーマットで、2つの生物サンプルでの遺伝子のセットの示差発現を分析する。2つのサンプルのプローブを調製して、異なる標識成分で標識化する。2つの標識されたプローブの混合液をアレイエレメントにハイブリダイズし、2つの異なる標識からの発光を個々に検出可能な条件下で2つの異なるシグナルを調べる。両方の生物サンプルから由来した同数のプローブにハイブリダイズするアレイ上のエレメントは、特有の結合した蛍光を出す (Shalton WO95/35505)。

【0103】

ハイブリダイゼーション複合体は、Cy3の励起のためには488nm、Cy5の励起のためには632nmでスペクトル線を生じ得るINNOVA70混合ガス10Wレーザー (Coherent, Santa Clara CA)を備えた顕微鏡で検出する。20x顕微鏡対物レンズ (Nikon, Melville NY)を用いて、アレイ上に励起レーザー光の焦点を当てる。アレイを含むスライドを顕微鏡のコンピュータ制御のX-Yステージに置き、20 μ mの解像度で対物レンズを通過してラスタースキャンする。示差ハイブリダイゼーションフォーマットで、レーザーにより2つの蛍光色素を同時に励起する。発光された光は、波長に基づき分離され、2つの蛍光色素に対応する2つの光電子増倍管検出器 (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ)に送られる。アレイと光電子増倍管間に設置された好適なフィルターを用いて、シグナルをフィルターする。用いる蛍光色素の最大発光の波長は、Cy3では565nm、Cy5では650nmである。プローブ混合液に加えた酵母対照mRNAによって生み出されるシグナル強度を用いてスキャンの感度を校正する。アレイ上の特定の位置には相補的DNA配列を含め、その位置におけるシグナルの強度がハイブリダイズする種の重量比1:100,000に相関するようにする。

【0104】

光電子増倍管の出力は、IBMコンパチブルPCコンピュータにインストールされた12ビットRTI-835Hアナログ-デジタル (A/D)変換ボード (Analog Devices, Inc., Norwood MA)を用いてデジタル化される。ディ

10

20

30

40

50

デジタル化されたデータは、青色（低シグナル）から赤色（高シグナル）までの擬似カラー範囲へのリニア20色変換を用いてシグナル強度がマッピングされたようなイメージとして表示される。データは、定量的にも分析される。2つの異なる蛍光色素を同時に励起及び測定する場合には、各蛍光色素の発光スペクトルを用いて、データは先ず蛍光色素間の光学的クロストーク（発光スペクトルの重なり起因する）を補正する。グリッドが蛍光シグナルイメージ上に重ねられ、それによって各スポットからのシグナルはグリッドの各エレメントに集められる。各エレメント内の蛍光シグナルは統合され、シグナルの平均強度に応じた数値が得られる。シグナル分析に用いるソフトウェアは、GEMTOOLSプログラム（Incycyte Genomics）である。

【0105】

10

9 相補的分子

cDNAに相補的な分子、約5（PNA）から5000 bp（cDNAインサートの相補体）を用いて、遺伝子発現を検出あるいは阻害する。LASERGENEソフトウェア（DNASTAR）を用いて、これらの分子を選択する。検出については、実施例7に記載されている。プロモーターが結合するのを妨げることにより転写を阻害するため、相補的分子を最も固有な5'配列に結合し、オープンリーディングフレームの開始コドンの5' UTR上流のヌクレオチドを含むよう設計する。相補的分子はゲノム配列（エンハンサーまたはイントロン等）を含み、ポリメラーゼ、転写因子または調節分子の結合のために十分に開くような二重らせんの能力に影響を及ぼす「三重らせん」塩基対の形成に用いられる。翻訳を阻害するためには、相補的分子を設計して、リボソームがタンパク質をコードするmRNAに結合するのを阻害する。

20

【0106】

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

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

30

【0107】

10 タンパク質発現

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

40

【0108】

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

【0109】

11 抗体の産生

ポリアクリルアミドゲル電気泳動法を用いて腫瘍抗原を精製して、マウスやウサギを免疫

50

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

【0110】

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

【0111】

12 特異的抗体を用いる天然タンパク質の精製

タンパク質を特異結合する抗体を用いて、イムノアフィニティークロマトグラフィにより天然あるいは組換えタンパク質を精製する。イムノアフィニティークラムは、CNBr-活性化SEPHAROSE樹脂(APB)に抗体を共有結合させることにより形成する。タンパク質を含む培養液をイムノアフィニティークラムに通し、タンパク質を優先的に吸着する条件下(例えば洗浄剤が存在する高イオン強度緩衝液)でカラムを洗浄する。結合後、そのタンパク質を、抗体とタンパク質との結合を切るために、例えば、pH2~3のバッファー、或いは高濃度の尿素またはチオシアン酸塩イオンを用いてカラムから溶出させ、タンパク質を回収する。

【0112】

13 cDNAまたはタンパク質と特異結合するためのスクリーニング分子

cDNAまたはその断片、タンパク質またはその部分を³²P-dCTP、Cy3-dCTPまたはCy5-dCTP(APB)あるいはBIODIPYかFITC(Molecular Probes, Eugene OR)で標識化する。前もって基質上に配置した候補分子または複合体のライブラリを標識したcDNAまたはタンパク質の存在下でインキュベートする。核酸配列かアミノ酸配列のための条件下でインキュベート後、基質を洗浄し、標識を保持する基質上の、特異結合か複合体成形を示す任意の位置がアッセイし、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られたデータを使用して、標識された核酸かタンパク質と結合した分子の間の親和性を計算する。

【0113】

14 2つのハイブリッドスクリーン

酵母2ハイブリッドシステム、MATCHMAKER LexA 2ハイブリッドシステム(Clontech Laboratories, Palo Alto CA)を使用して、本発明のタンパク質を結合するペプチドをスクリーニングする。タンパク質をコードするcDNAをpLexAベクター、リガンドの多数のクローニング部位に挿入して、大腸菌に形質転換する。mRNAから調製したcDNAをpB42ADベクターの多数のクローニング部位に挿入して、ライゲーションし、cDNAライブラリを作製するために大腸菌に形質転換する。pLexAプラスミドとpB42AD-cDNAライブラリ作製を大腸菌から単離し、ポリエチレングリコール/リチウムアセテートプロトコルを用いて形質転換受容性をもつ酵母EGY48[p8op-lacZ]細胞に同時形質転換するために比率2:1で使用する。形質転換した酵母細菌を、ヒスチジン(-His)、トリプトファン(-Trp)、ウラシル(-Ura)を含まない合成ドロップアウト(SD)培地で培養する、そしてコロニーが成長して数えられるまで、30でインキュベートする。コ

コロニ-を最小限の容積の1 x T E (p H 7 . 5) で集め、2 % ガラクトース (G a l) , 1 % ラフィノース (R a f) 、 8 0 m g / m l 5 - ブロモ - 4 - クロロ - 3 - インドリル - d - ガラクトピラノシド (X - G a l) で補われる S D / - H i s / - L e u / - T r p / - U r a 培地で再培養する。次いで青色コロニ-の成長を検査する。発現したタンパク質とc D N A 融合タンパク質間の相互作用により、E G Y 4 8 の L E U 2 レポーター遺伝子を活性化し、ロイシン (- L e u) を含まない培地でコロニ-成長を起こす。相互作用はまた、X - G a l で成長するコロニ-に青色を産生する p 8 o p - l a c Z レポーター作成物からの - ガラクトシダーゼの発現を活性化する。

【 0 1 1 4 】

発現したタンパク質とc D N A 融合タンパク質間の陽性の相互作用が、個々の陽性コロニ-を単離して、3 0 で1から2日間、S D / - T r p / - U r a 液体で育成することにより確認できる。培地のサンプルをS D / - T r p / - U r a 培地で培養し、コロニ-が出現するまで3 0 でインキュベートする。サンプルをS D / - T r p / - U r a と S D / - H i s / - T r p / - U r a プレートで複製培養する。ヒスチジンを含まない培地では育成せず、ヒスチジンを含むS D で育成するコロニ-は、p L e x A プラスミドを失っている。ヒスチジンを必要とするコロニ-をS D / G a l / R a f / X - G a l / - T r p / - U r で育成する。そして白色コロニ-を単離して、増殖する。タンパク質と物理的に相互作用するタンパク質をコードするc D N A p B 4 2 A D - c D N A プラスミドを酵母細胞から単離して、特徴付ける。

【 0 1 1 5 】

1 5 転写イメージング

L I F E S E Q G O L D データベース (J u n 0 1 r e l e a s e , I n c y t e G e n o m i c s) を用いて転写イメージを実施した。このプロセスにより、1 4 0 0 を超えるc D N A ライブラリで発現したc D N A の相対存在量の評価が可能となる。転写イメージングの基準をカテゴリー、ライブラリ毎のc D N A の数、ライブラリ説明、疾患の兆候、サンプルの臨床の関連性等から選択することが可能である。

【 0 1 1 6 】

L I F E S E Q データベースのすべての配列とc D N A ライブラリを体系、器官/組織、細胞のタイプにより分類した。各カテゴリーでは、配列が発現したライブラリ数を数えて、そのカテゴリー内のライブラリの総数を示した。ある転写イメージでは、すべての規準化またはサブトラクションしたライブラリ (プロセシングの前に多数のコピー配列が除かれている) 、およびすべての混合またはプールした組織 (2 以上の組織の種類または2 以上の被験者の組織を含有する点でと非特異的であると考慮される) を分析から排除した。臨床の関連性が強調される場合、処理または非処理細胞株と/あるいは胎児組織データも無視または除去することができる。逆に分析から臨床サンプルを除去することにより、遺伝疾患または幹細胞からの特定の細胞あるいは器官 (神経、心臓または腎臓) の分化の解明が促進される場合は胎児組織を強調することもできる。

【 0 1 1 7 】

S E Q I D N O : 1 , 5 , 1 0 の転写イメージを下記に示す。列1はライブラリ名、列2はライブラリで配列化したc D N A の数、列3はライブラリ説明、列4はライブラリの転写の絶対存在量、列5はライブラリの転写の存在量%を示す。

【 外 5 】

カテゴリー全て (SEQ ID NO:1)

ライブラリ* cDNA	前立腺組織の説明	存在量	存在量%
CONDUT01 1286	腹膜、神経内分泌癌、66F	2	0.1555
PENHTUE02 1846	陰茎、扁平上皮細胞癌、64M、5RP	1	0.0542
LUNGTUT09 3969	肺 扁平上皮細胞癌、68M	2	0.0504
OVARTUM02 2932	卵巣 乳頭 漿液 癌、64F、WM/WN	1	0.0341
SPLNTUT02 3077	脾臓 ホジキン病、45M	1	0.0325
COLITUT02 6656	回盲、バーキットリンパ腫、29F	2	0.0300

*細胞株ライブラリ、胎児ライブラリ、プールしたライブラリ、規準化ライブラリ、またはサブストラクションしたライブラリは、この分析では用いられなかった。

10

【0118】

腹膜の神経内分泌腺癌のSEQ ID NO:1の示差発現は、消化管の他のすべての組織の発現よりも存在量%が3倍高い。細胞学的に正常な組織では、発現は見られなかった。細胞または組織に特異な診断手順で用いられ、確立した標準と比較する場合、SEQ ID NO:1は癌、特に腹膜の神経内分泌癌の診断に有用である。

【外6】

カテゴリー外分泌腺 (胸)

ライブラリ* cDNA	膀胱組織の説明	存在量	存在量%
BRSTUNF01 1146	乳房腫瘍系統 T-47D、管癌、54F	1	0.0873
BRSTTUT16 3724	乳房管癌、43F、m/BRSTTMT01	2	0.0537
BRSTTUT08 3928	乳房腫瘍、腺癌、45F、m/BRSTNOT09	2	0.0509
BRSTUNT01 3130	乳房腫瘍系統 T47D、54F	1	0.0319
BRSTNOT03 6777	BRSTTUT02 管 腺癌に一致、54F	1	0.0148
BRSTTUT13 7631	乳房腺癌、46F、m/BRSTNOT33	1	0.0131
BRSTTUT03 10092	乳房 小葉癌、58F、m/BRSTNOT05	1	0.0099

*この分析で排除されたライブラリはない。

20

【0119】

SEQ ID NO:5は、乳房腫瘍系統T-47Dと癌性と正常な乳房組織の一致するセットの発現で示されるように乳癌を診断するのに有用である。乳房容量減少手術中に患者から切除した細胞学的に正常な乳房組織、または他のいかなる乳房ライブラリにも発現は見られなかった。乳房組織に関して使用する場合には、SEQ ID NO:1は乳癌の診断に有用である。

30

【外7】

カテゴリー消化管 (結腸)

ライブラリ cDNA	肺組織の説明	存在量	存在量%
COLNTUP12 2312	結腸腺癌、M/F、プール、3' CGAP	1	0.0433
COLNTUP15 12065	結腸腺癌、プール、NORM、3' CGAP	5	0.0414
COLNTUN03 2462	結腸腺癌、M/F、プール、NORM	1	0.0406
COLNTUP17 7421	結腸腺癌、3'、CGAP	2	0.0270
COLITUT02 6656	バーキットリンパ腫、29F、m/COLANOT03	1	0.0150
COLNTUP16 8499	結腸腺癌、プール、NORM、3'/5' CGAP	1	0.0118

40

【0120】

SEQ ID NO:10の示差発現は、潰瘍性大腸炎 (COLADIT05、COLANOT02、COLAUCT01、COLDDIE01)、良性家族性ポリープ症 (COLCDIT01、COLDNOT01、COLTDIT04)、潰瘍性大腸炎 (COLNDIP02、COLNNOT23、COLNUCT03、COLSUCT01) または細胞学的に正常な組織 (COLNNO5、COLNNO1、COLNNO2、COLNNO1、COLNNO5、COLNNO7、COLNNO8、COLNNO9、COLNNO11、COLNNO13、COLNNO16、

50

COLNNOT19、COLNNOT22)と診断された患者の組織から作製したライブラリに見られなかった。細胞または組織に特異な診断手順で用いられ、確立した標準と比較する場合、SEQ ID NO: 1は結腸癌の診断に有用である。確立した標準と患者サンプルを使用するアッセイでは、このタンパク質に特異的に結合するcDNA、mRNA、タンパク質または抗体は、細胞周期疾患のための臨床的に意義のある診断マ - カ - として役立つ。

【0121】

本明細書において記載した全ての特許出願、刊行物は、言及することをもって本明細書の一部とする。当業者は、本発明の範囲及び精神から逸脱することなく本発明の記載した方法及びシステムの種々の改変を行い得る。本発明について説明するにあたり特定の好適実施例に関連して説明を行ったが、本発明の範囲が、そのような特定の実施例に不当に制限されるべきではないことを理解されたい。実際に、分子生物学または関連分野の専門家には明らかな、本明細書に記載されている本発明の実施方法の様々な改変は、特許請求の範囲内にあるものとする。

【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
7 March 2002 (07.03.2002)

PCT

(10) International Publication Number
WO 02/18575 A2

- (51) International Patent Classification: C12N 15/00
- (21) International Application Number: PCT/US01/26682
- (22) International Filing Date: 27 August 2001 (27.08.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/229,253 30 August 2000 (30.08.2000) US
- (71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (72) Inventors: and
(75) Inventors/Applicants (for US only): WALKER, Michael, C. [CA/US]; 1050 Boregas Avenue, #80, Sunnyvale, CA 94089 (US); JUNG, Kenneth [US/US]; 725 Van Ness Avenue, #203, San Francisco, CA 94102 (US).
- (74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 02/18575 A2

(54) Title: GENES EXPRESSED IN THE CELL CYCLE

(57) Abstract: The invention provides cDNAs, their encoded proteins, and antibodies which may be used in methods for diagnosing and treating cell cycle disorders.

WO 02/18575

PCT/US01/26682

GENES EXPRESSED IN THE CELL CYCLE**TECHNICAL FIELD**

The invention relates to cDNAs identified by their co-expression with known cell cycle genes and to their use in diagnosis, prognosis, treatment, and evaluation of therapies for cell cycle disorders.

BACKGROUND OF THE INVENTION

Cell division is the fundamental process by which all living things grow, repair, and reproduce. In unicellular organisms, each cell division doubles the number of organisms; and in multicellular species, many rounds of cell division are required to produce a new organism or to replace cells lost by wear and tear or by programmed cell death. Details of the cell division cycle vary, but the basic process consists of three principle events. The first event, interphase, involves preparation for cell division, replication of the DNA, and production of essential proteins. In the second event, mitosis, the nuclear material is divided and separates to opposite sides of the cell. The final event, cytokinesis, is division of the cytoplasm. The sequence and timing of cell cycle events is under the control of cell cycle regulators which control the process by positive or negative mechanisms at various check points.

Cancers and immune conditions, diseases and disorders are associated with the dysregulation of normal cell proliferation. In cancer, this dysregulation is often attributable to oncogenes, mutant isoforms of normal cellular genes. In some cases, these oncogenes are activated by viruses as a consequence of the integration of a viral genome into the DNA of the host cell. Sometimes, more than one oncogene, capable of maintaining the infected cell in a condition of continuous cell division, is activated. Other oncogenes are abnormally expressed with respect to location or level of expression. This latter category causes cancer by altering transcriptional control of cell proliferation. At least five classes of oncogenes are known; they include cytokines and growth factors; receptors such as *erbA*, *erbB*, *neu*, and *ros*; intracellular signal transducers such as *src*, *yes*, *fps*, *abl*, and *met*; nuclear transcription factors such as *fos*; cell-cycle control proteins such as *RB* and *p53*; and mutated tumor-suppressor genes such as *mdm2*, *sec*, and *ras* (Bohmann *et al.* (1987) *Science* 238:1386-1392; Cohen and Curran (1988) *Mol Cell Biol* 8:2063-2069; and van Straaten *et al.* (1983) *Proc Natl Acad Sci* 80:3183-3187).

For example, in cancer, oncogenes contribute to unrestricted cell proliferation through their involvement in the reception and transduction of growth factor signals and in the modulation of gene expression in response to these signals. Stimulation of a cell by growth factor activates two sets of genes, the early-response genes and the delayed-response genes. Early-response genes include the *myc*, *fos*, and *jun* proto-oncogenes, all of which encode gene regulatory proteins. These regulatory proteins activate the transcription of the delayed-response genes which encode proteins such as the

WO 02/18575

PCT/US01/26682

cyclins and cyclin-dependent kinases directly involved in cell cycle progression.

The discovery of cDNAs which coexpress with known cell cycle genes satisfies a need in the art by providing new compositions which are useful in the diagnosis, prognosis, treatment, and evaluation of therapies for cell cycle disorders.

5

SUMMARY OF THE INVENTION

The invention provides a composition comprising a plurality of cDNAs having the nucleic acid sequences of SEQ ID NOs: 1-10 or their complements that are coexpressed with one or more known cell cycle genes in a plurality of biological samples. The invention also provides a method of using a composition to screen a plurality of molecules to identify at least one ligand which specifically binds a cDNA of the composition, the method comprising combining the composition with molecules under conditions to allow specific binding; and detecting specific binding, thereby identifying a ligand which specifically binds the cDNA. In one embodiment, the molecules are selected from DNA molecules, RNA molecules, peptide nucleic acids, transcription factors, enhancers, repressors, mimetics, and proteins.

15

The invention provides a method for using a composition to detect gene expression in a sample containing nucleic acids, the method comprising hybridizing the composition to the nucleic acids under conditions for formation of one or more hybridization complexes; and detecting hybridization complex formation, wherein complex formation indicates gene expression in the sample. In one embodiment, the cDNAs of the composition are attached to a substrate. In another embodiment, complex formation when compared to standards is diagnostic of cell cycle disorders.

20

The invention provides an isolated cDNA having a nucleic acid sequence selected from SEQ ID NOs: 1, 2, and 4-10 and the complements thereof. In different aspects, each cDNA is used as a diagnostic, as a probe, in an expression vector, and in assessing the prognosis and treatment of a cell cycle disorder. The invention also provides a composition comprising a cDNA and a labeling moiety.

25

The invention further provides a method for using a cDNA to screen a plurality of molecules to identify a ligand which specifically binds the cDNA, the method comprising combining the cDNA with a sample under conditions to allow specific binding; recovering the bound cDNA; and separating the ligand from the bound cDNA, thereby obtaining purified ligand. In one embodiment, the molecules to be screened are selected from DNA molecules, RNA molecules, peptide nucleic acids, transcription factors, enhancers, repressors, mimetics, and proteins.

30

The invention provides a method for using a cDNA to detect gene expression in a sample containing nucleic acids, the method comprising hybridizing the cDNA to nucleic acids of a sample under conditions for formation of one or more hybridization complexes; and detecting hybridization complex formation, wherein complex formation indicates gene expression in the sample. In one

35

embodiment, the cDNA is attached to a substrate. In another embodiment, gene expression when

WO 02/18575

PCT/US01/26682

compared to standards is diagnostic of a cell cycle disorder. The method also provides a vector containing the cDNA, a host cell containing a vector and a method for using a host cell to produce a protein or peptide encoded by the cDNA comprising culturing the host cell under conditions for expression of the protein; and recovering the protein from cell culture.

5 The invention provides a purified protein encoded by a cDNA of the invention. The invention also provides a method for using the protein or peptide to screen a plurality of molecules to identify and purify a ligand which specifically binds the protein. In one embodiment, the molecules to be screened are selected from DNA molecules, RNA molecules, peptide nucleic acids, proteins, agonists, antagonists, and antibodies.

10 The invention provides a method of using a protein to prepare and purify antibodies comprising immunizing an animal with the protein or peptide 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; and dissociating the antibodies from the protein, thereby obtaining purified antibodies. The invention also provides
15 methods for using an antibody which specifically binds the protein to diagnose a cell cycle disorder, the method comprising combining an antibody with a sample under conditions for specific binding, detecting antibody complex formation, comparing antibody complex formation with a standard, thereby diagnosing a cell cycle disorder. The invention further provides a composition comprising a cDNA, a protein or an antibody that specifically binds a protein or peptide and a pharmaceutical
20 carrier for use in treating a cell cycle disorder.

DESCRIPTION OF THE INVENTION

It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include the plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an
25 antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

DEFINITIONS

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

35 "Cell cycle gene" refers to a cDNA which has been previously identified as useful in the

WO 02/18575

PCT/US01/26682

diagnosis, prognosis, treatment, and evaluation of therapies associated with unregulated cell cycling. Typically, this means that the known gene is differentially expressed at higher (or lower) levels in tissues from patients with a cell cycle disorder when compared with normal expression in any tissue. The cell cycle genes used in this invention and described in EXAMPLE IV are *cdc2*, *cdc7*, *cdc23*,
5 *cyclin B*, *hBub1*, *HKSP*, *hp55cdc*, *MCAK*, *mitosin*, *mki67a*, *MKLP-1*, *myb*, *nlk1*, *cdc21*, *PRC1*, *Aik2*, *survivin*, *topoII*, and *UbcH10*.

"Cell cycle disorder" refers to any cancer or immune disorder including, but not limited to, an adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma or cancers of the blood, bone, bone marrow, brain, breast, gastrointestinal tract (esophagus, stomach, small intestine or colon),
10 heart, kidney, liver, lung, lymph, muscle, nerve, ovary, pancreas, prostate, skin, spleen, testis, and uterus; asthma, atherosclerosis, Crohn's disease, glomerulonephritis, multiple sclerosis, myasthenia gravis, osteoporosis, rheumatoid arthritis, scleroderma, and systemic lupus erythematosus.

"cDNA" refers to an isolated polynucleotide or any fragment or oligonucleotide thereof. It may of genomic or synthetic origin, double-stranded or single-stranded, and combined with
15 carbohydrate, lipids, protein or other materials to perform a particular activity or form a useful composition.

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

20 "Isolated or purified" refers to a cDNA or protein that is removed from its natural environment and that is separated from other components with which it is naturally present.

"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
25 minerals, cofactors, nucleic acids, proteins, carbohydrates, fats, and lipids.

"Protein" refers to a polypeptide, or any portion or oligopeptide thereof whether naturally occurring or synthetic.

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

"Similarity" refers to the quantification (usually percentage) of nucleotide or residue matches between at least two sequences aligned using a standard algorithm such as Smith-Waterman
35 alignment (Smith and Waterman (1981) *J Mol Biol* 147:195-197) or BLAST2 (Altschul *et al.* (1997)

WO 02/18575

PCT/US01/26682

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

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

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

A "transcript image" is a profile of gene transcription activity in a particular tissue at a
15 particular time.

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

THE INVENTION

The present invention utilizes a method for identifying cDNAs or proteins that are associated
25 with a specific disease, regulatory pathway, subcellular compartment, cell type, tissue type, or species. In particular, the method identifies cDNAs useful in diagnosis, prognosis, treatment, and evaluation of therapies for cell cycle disorders.

The method provides for the identification of cDNAs that are expressed in a plurality of libraries. The expression patterns of genes with known function are compared with those of cDNAs
30 with unknown function to determine whether a specified co-expression probability threshold is met. Through this comparison, a subset of the cDNAs having a high co-expression probability with the known genes can be identified.

The cDNAs originate from cDNA libraries derived from a variety of sources including, but not limited to, eukaryotes such as human, mouse, rat, dog, monkey, plant, and yeast; prokaryotes such
35 as bacteria; and viruses. These cDNAs can also be selected from a variety of sequence types

WO 02/18575

PCT/US01/26682

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.

5 The cDNA libraries used in the co-expression analysis can be obtained from adrenal gland, biliary tract, bladder, blood cells, blood vessels, bone marrow, brain, bronchus, cartilage, chromaffin system, colon, connective tissue, cultured cells, embryonic stem cells, endocrine glands, epithelium, esophagus, fetus, ganglia, heart, hypothalamus, immune system, intestine, islets of Langerhans, kidney, larynx, liver, lung, lymph, muscles, neurons, ovary, pancreas, penis, peripheral nervous
10 system, peritoneum, phagocytes, pituitary, placenta, pleurus, prostate, salivary glands, seminal vesicles, skeleton, spleen, stomach, testis, thymus, tongue, ureter, uterus, and the like. The number of cDNA libraries selected can range from as few as 5 to greater than 10,000. Preferably, the number of the cDNA libraries is greater than 500.

In a preferred embodiment, the cDNAs are assembled from related sequences, such as
15 sequence fragments derived from a single transcript. Assembly of the polynucleotide can be performed using sequences of various types including, but not limited to, ESTs, extension of the EST, shotgun sequences from a cloned insert, or full length cDNAs. In a most preferred embodiment, the cDNAs are derived from human sequences that have been assembled using the algorithm disclosed in USSN 9,276,534, filed March 25, 1999, incorporated herein by reference.

20 Experimentally, differential expression of the cDNAs can be evaluated by methods including, but not limited to, differential display by spatial immobilization or by gel electrophoresis, genome mismatch scanning, representational difference analysis, and transcript imaging. Representative transcript images for SEQ ID NO:s 1, 5 and 10 are found in EXAMPLE XV. The transcript images confirm the data produced by the co-expression method disclosed herein. Additionally, differential
25 expression can be assessed by microarray technology. Any of these methods may be used alone or in combination.

Known cell cycle genes can be selected based on function and the use of the genes as diagnostic or prognostic markers or as therapeutic targets for diseases associated with unregulated cell proliferation. Preferably, the known cell cycle genes include *cdc2*, *cdc7*, *cdc23*, *cyclin B*, *hBub1*,
30 *HKSP*, *hp55cdc*, *MCAK*, *mitosin*, *mki67a*, *MKLP-1*, *myb*, *nlk1*, *cdc21*, *PRC1*, *Aik2*, *survivin*, *topoII*, and *UbcH10*.

The procedure for identifying cDNAs that exhibit a statistically significant co-expression pattern with known cell cycle genes is as follows. First, the presence or absence of a gene sequence in a cDNA library is defined: a gene is present in a cDNA library when at least one cDNA fragment
35 corresponding to that gene is detected in a cDNA sample taken from the library, and a gene is absent

WO 02/18575

PCT/US01/26682

from a library when no corresponding cDNA fragment is detected in the sample.

Second, the significance of gene co-expression is evaluated using a probability method to measure a due-to-chance probability of the co-expression. The probability method can be the Fisher exact test, the chi-squared test, or the kappa test. These tests and examples of their applications are well known in the art and can be found in standard statistics texts (Agresti (1990) Categorical Data Analysis, John Wiley & Sons, New York NY; Rice (1988) Mathematical Statistics and Data Analysis, Duxbury Press, Pacific Grove CA). A Bonferroni correction (Rice, *supra*, p. 384) can also be applied in combination with one of the probability methods for correcting statistical results of one gene versus multiple other genes. In a preferred embodiment, the due-to-chance probability is measured by a Fisher exact test, and the threshold of the due-to-chance probability is set preferably to less than 0.001, more preferably to less than 0.00001.

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

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

For a given pair of genes, the co-occurrence data is summarized in a 2 x 2 contingency table (below).

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

The contingency table shows the co-occurrence data for gene A and gene B in a total of 30 libraries. Both gene A and gene B occur 10 times in the libraries, and the table summarizes and presents: 1) the number of times gene A and B are both present in a library; 2) the number of times gene A and B are both absent in a library; 3) the number of times gene A is present, and gene B is absent; and 4) the number of times gene B is present, and gene A is absent. The upper left entry is the number of times the two genes co-occur in a library, and the middle right entry is the number of times neither gene occurs in a library. The off diagonal entries are the number of times one gene occurs, and the other does not. Both A and B are present eight times and absent 18 times. Gene A is present, and gene B is absent, two times; and gene B is present, and gene A is absent, two times. The probability ("p-value") that the above association occurs due to chance as calculated using a Fisher exact test is 0.0003. Associations are generally considered significant if a p-value is less than 0.01

WO 02/18575

PCT/US01/26682

(Agestri, *supra*; Rice, *supra*).

This method of estimating the probability for co-expression of two genes makes several assumptions. The method assumes that the libraries are independent and are identically sampled. However, in practical situations, the selected cDNA libraries are not entirely independent, because more than one library may be obtained from a single subject or tissue. Nor are they entirely identically sampled, because different numbers of cDNAs may be sequenced from each library. The number of cDNAs sequenced typically ranges from 5,000 to 10,000 cDNAs per library. In addition, because a Fisher exact co-expression probability is calculated for each gene versus 37,071 other assembled genes that occur in at least five libraries, a Bonferroni correction for multiple statistical tests is used.

Using the method above, we have identified cDNAs that exhibit strong association, or co-expression, with known genes that are specific to the cell cycle. The results presented in the co-expression table seen in EXAMPLE V are summarized in the table below. Column 1 is the SEQ ID number, column 2, the known cell cycle gene(s) with which the cDNA is most highly co-expressed; column 3, the p-value; and column 4, a cell cycle disorder for which the co-expressed cDNA is a specific diagnostic marker.

SEQ ID	Cell Cycle Gene	p-value	Cell Cycle Disorder
1	topo II	16	peritoneal neuroendocrine carcinoid
2	PRC1	12	colon adenocarcinoma
3	CDC23	12	lymphoma
4	topo II, PRC1	10	metastatic melanoma
5	cyclin B, Ubch10	13	breast cancer
6	PRC1	16	colon adenocarcinoma
7	cyclin B	9.5	brain cancer
8	topo II	13	testicular adenocarcinoma
9	topo II	9	metastatic melanoma
10	hp55cdc	17	colon adenocarcinoma

This table shows that the cDNAs claimed herein have a very highly significant co-expression (less than .0000001) with known cell cycle genes. Therefore, the cDNAs are useful as surrogate markers in diagnosis, prognosis, and evaluation of therapies for cell cycle disorders and potentially serve as therapeutics for the elimination or control of unregulated cell cycling. Further, the proteins or peptides expressed from the cDNAs are either potential therapeutics or targets for the identification or development of therapeutics. Similarly, antibodies made from or identified using the protein are either potential therapeutics or pharmaceutical carriers.

Therefore, in one embodiment, the present invention encompasses a composition of cDNAs comprising the nucleic acid sequences of SEQ ID NOs: 1-10 or the complements thereof. These ten cDNAs are shown by the method of the present invention to have strong co-expression with known cell cycle genes and with each other. The invention also provides a cDNA, its complement, and a

WO 02/18575

PCT/US01/26682

probe comprising the cDNA selected from SEQ ID NOs:1, 2, and 4-10. Variants typically have at least about 70% nucleic acid sequence identity to at least one of these sequences.

The cDNA or the encoded protein may be used to search against the GenBank primate (pri), rodent (rod), mammalian (mam), vertebrate (vrtp), and eukaryote (eukp) databases, SwissProt, BLOCKS (Bairoch *et al.* (1997) *Nucleic Acids Res* 25:217-221), PFAM, and other databases that contain previously identified and annotated motifs, sequences, and gene functions. Methods that search for primary sequence patterns with secondary structure gap penalties (Smith *et al.* (1992) *Protein Engineering* 5:35-51) as well as algorithms such as Basic Local Alignment Search Tool (BLAST; Altschul (1993) *J Mol Evol* 36:290-300; Altschul *et al.* (1990) *J Mol Biol* 215:403-410), BLOCKS (Henikoff and Henikoff (1991) *Nucleic Acids Res* 19:6565-6572), Hidden Markov Models (HMM; Eddy (1996) *Cur Opin Str Biol* 6:361-365; Sonnhammer *et al.* (1997) *Proteins* 28:405-420), and the like, can be used to manipulate and analyze nucleotide and amino acid sequences. These databases, algorithms and other methods are well known in the art and are 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, p 856-853).

Also encompassed by the invention are polynucleotides that are capable of hybridizing to SEQ ID NOs:1-10, and fragments thereof under stringent conditions. Stringent conditions can be defined by salt concentration, temperature, and other chemicals and conditions well known in the art. Conditions can be selected, for example, by varying the concentrations of salt in the prehybridization, hybridization, and wash solutions or by varying the hybridization and wash temperatures. With some substrates, the temperature can be decreased by adding formamide to the prehybridization and hybridization solutions.

Hybridization can be performed at low stringency, with buffers such as 5xSSC (sodium saline citrate) with 1% sodium dodecyl sulfate (SDS) at 60°C, which permits complex formation between two 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 45°C (medium stringency) or 68°C (high stringency), to maintain hybridization of only those complexes that contain completely complementary sequences. Background signals can be reduced by the use of detergents such as SDS, sarcosyl, or TRITON X-100 (Sigma-Aldrich, St. Louis MO), and/or a blocking agent, such as salmon sperm DNA. Hybridization methods are described in detail in Ausubel (*supra*, units 2.8-2.11, 3.18-3.19 and 4-6-4.9) and Sambrook *et al.* (1989; *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview NY)

A cDNA can be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences such as promoters and other regulatory elements. (See, e.g., Dieffenbach and Dveksler (1995) *PCR Primer, a Laboratory Manual*, Cold

WO 02/18575

PCT/US01/26682

Spring Harbor Press, Plainview NY). Additionally, one may use an XL-PCR kit (Applied Biosystems (ABI), Foster City CA), nested primers, and commercially available cDNA libraries (Life Technologies, Rockville MD) or genomic libraries (Clontech, Palo Alto CA) to extend the sequence. For all PCR-based methods, primers may be designed using commercially available software (LASERGENE software, DNASTAR, Madison WI) or another program, to be about 15 to 30 nucleotides in length, to have a GC content of about 50%, and to form a hybridization complex at temperatures of about 68°C to 72°C.

In another aspect of the invention, the cDNA can be cloned into a recombinant vector that directs the expression of the protein, or structural or functional portions thereof, in host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode the same or a functionally equivalent amino acid sequence may be produced and used to express the protein encoded by the cDNA. The nucleotide sequences can be engineered using methods generally known in the art in order to alter the nucleotide sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In order to express a biologically active protein, the cDNA or derivatives thereof, may be inserted into an expression vector, i.e., a vector which contains the elements for transcriptional and translational control of the inserted coding sequence in a particular host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions. Methods which are well known to those skilled in the art may be used to construct such expression vectors. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook, *supra*; Ausubel, *supra*).

A variety of expression vector/host cell systems may be utilized to express the cDNA. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with baculovirus vectors; plant cell systems transformed with viral or bacterial expression vectors; or animal cell systems. For long term production of recombinant proteins in mammalian systems, stable expression in cell lines is preferred. For example, the cDNA can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable or visible marker gene on the same or on a separate vector. The invention is not to be limited by the vector or host cell employed.

WO 02/18575

PCT/US01/26682

In general, host cells that contain the cDNA and that express the protein may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or amino acid sequences. Immunological methods for detecting and measuring the expression of the protein using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS).

Host cells transformed with the cDNA may be cultured under conditions for the expression and recovery of the protein from cell culture. The protein produced by a transgenic cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing the cDNA may be designed to contain signal sequences which direct secretion of the protein through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the protein include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the ATCC (Manassas VA) and may be chosen to ensure the correct modification and processing of the expressed protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences are ligated to a heterologous sequence resulting in translation of a fusion protein containing heterologous protein moieties in any of the aforementioned host systems. Such heterologous protein moieties facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase, maltose binding protein, thioredoxin, calmodulin binding peptide, 6-His, FLAG, c-myc, hemagglutinin, and monoclonal antibody epitopes.

In another embodiment, the cDNAs, wholly or in part, are synthesized using chemical or enzymatic methods well known in the art (Caruthers *et al.* (1980) Nucl Acids Symp Ser (7) 215-233; Ausubel, *supra*). For example, peptide synthesis can be performed using various solid-phase techniques (Roberge *et al.* (1995) Science 269:202-204), and machines such as the ABI 431A peptide synthesizer (ABI) can be used to automate synthesis. If desired, the amino acid sequence may be altered during synthesis and/or combined with sequences from other proteins to produce a variant.

WO 02/18575

PCT/US01/26682

SCREENING, DIAGNOSTICS AND THERAPEUTICS

The compositions or cDNAs can be used in diagnosis, prognosis, treatment, and selection and evaluation of therapies for cell cycle disorders including, but not limited to, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma or cancers of the blood, bone, bone marrow, brain, breast, gastrointestinal tract (esophagus, stomach, small intestine or colon), heart, kidney, liver, lung, lymph, muscle, nerve, ovary, pancreas, prostate, skin, spleen, testis, and uterus; asthma, atherosclerosis, Crohn's disease, glomerulonephritis, multiple sclerosis, myasthenia gravis, osteoporosis, rheumatoid arthritis, scleroderma, and systemic lupus erythematosus.

The compositions or cDNAs may be used to screen a plurality of molecules for specific binding affinity. The assay can be used to screen a plurality of DNA molecules, RNA molecules, peptide nucleic acids (PNAs), peptides, ribozymes, antibodies, agonists, antagonists, immunoglobulins, inhibitors, proteins including transcription factors, enhancers, repressors, and drugs and the like which regulate the activity of the polynucleotide in the biological system. The assay involves providing a plurality of molecules, combining the cDNA or a fragment thereof with the plurality of molecules under conditions suitable to allow specific binding, and detecting specific binding to identify at least one molecule which specifically binds the cDNA.

Similarly the proteins or portions thereof may be used to screen libraries of molecules or compounds in any of a variety of screening assays. The portion of a 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. Specific binding between the protein and the molecule may be measured. The assay can be used to screen a plurality of DNA molecules, RNA molecules, PNAs, peptides, mimetics, ribozymes, antibodies, agonists, antagonists, immunoglobulins, inhibitors, peptides, polypeptides, drugs and the like, which specifically bind the protein. One method for high throughput screening using very small assay volumes and very small amounts of test compound is described in Burbaum *et al.* USPN 5,876,946, incorporated herein by reference, which screens large numbers of molecules for enzyme inhibition or receptor binding.

In one preferred embodiment, the cDNAs are used for diagnostic purposes to determine the absence, presence, or altered--increased or decreased compared to a normal standard-- expression of the gene. The polynucleotide consists of complementary RNA and DNA molecules, branched nucleic acids, and/or PNAs. In one alternative, the cDNAs are used to detect and quantify gene expression in samples in which expression of the cDNA is correlated with disease. In another alternative, the cDNA can be used to detect genetic polymorphisms associated with a disease. These polymorphisms may be detected in the transcript cDNA.

The specificity of the probe is determined by whether it is made from a unique region, a regulatory region, or from a conserved motif. Both probe specificity and the stringency of diagnostic

WO 02/18575

PCT/US01/26682

hybridization or amplification (maximal, high, intermediate, or low) will determine whether the probe identifies only naturally occurring, exactly complementary sequences, allelic variants, or related sequences. Probes designed to detect related sequences should preferably have at least 50% sequence identity to any of the cDNAs.

5 Methods for producing hybridization probes include the cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by adding RNA polymerases and labeled nucleotides. Hybridization probes may incorporate nucleotides labeled by a variety of reporter groups including, but not limited to, radionuclides such as ³²P or ³⁵S, enzymatic
10 labels such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, fluorescent labels, and the like. The labeled cDNAs may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; and in microarrays utilizing samples from subjects to detect altered protein expression.

The cDNAs can be labeled by standard methods and added to a sample from a subject under
15 conditions for the formation and detection of hybridization complexes. After incubation the sample is washed, and the signal associated with hybrid complex formation is quantitated and compared with a standard value. Standard values are derived from any control sample, typically one that is free of the suspect disease. If the amount of signal in the subject sample is altered in comparison to the standard value, then the presence of altered levels of expression in the sample indicates the presence
20 of the disease. Qualitative and quantitative methods for comparing the hybridization complexes formed in subject samples with previously established standards are well known in the art.

Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual subject. Once the presence of disease is established and a treatment protocol is initiated, hybridization or
25 amplification assays can 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 healthy subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to many years.

The cDNAs may also be used on a microarray to monitor the expression patterns. The
30 microarray may also be used to identify splice variants, mutations, and polymorphisms. Information derived from analyses of the expression patterns may be used to determine gene function, to understand the genetic basis of a disease, to diagnose a disease, and to develop and monitor the activities of therapeutic agents used to treat a disease. Microarrays may also be used to detect genetic diversity, single nucleotide polymorphisms which may characterize a particular population, at the
35 genome level.

WO 02/18575

PCT/US01/26682

In yet another alternative, cDNAs may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Fluorescent *in situ* hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data as described in Heinz-Ulrich *et al.* (In: Meyers, *supra*, pp. 965-968).

5 In another embodiment, antibodies or Fabs comprising an antigen binding site that specifically binds the protein may be used for the diagnosis and prognosis of diseases characterized by the over-or-under expression of the protein. A variety of protocols for measuring protein expression, including ELISAs, RIAs, and FACS, are well known in the art and provide a basis for diagnosing altered or abnormal levels of expression. Standard values for protein expression are
10 established by combining samples taken from healthy subjects, preferably human, with antibody to the protein under conditions for complex formation. The amount of complex formation may be quantitated by various methods, preferably by photometric means. Quantities of the protein expressed in disease samples are compared with standard values. Deviation between standard and subject values establishes the parameters for diagnosing or monitoring disease. Alternatively, one
15 may use competitive drug screening assays in which neutralizing antibodies capable of binding specifically with the protein compete with a test compound. Antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with the protein. In one aspect, the antibodies can be used for treatment or monitoring therapeutic treatment for cell cycle disorders.

20 In another aspect, the cDNA, or its complement, may be used therapeutically for the purpose of expressing mRNA and protein, or conversely to block transcription or translation of the mRNA. Expression vectors may be constructed using elements from retroviruses, adenoviruses, herpes or vaccinia viruses, or bacterial plasmids, and the like. These vectors may be used for delivery of nucleotide sequences to a particular target organ, tissue, or cell population. Methods well known to
25 those skilled in the art can be used to construct vectors to express nucleic acid sequences or their complements. (See, e.g., Maulik *et al.* (1997) Molecular Biotechnology, Therapeutic Applications and Strategies, Wiley-Liss, New York NY.) Alternatively, the cDNA or its complement, may be used for somatic cell or stem cell gene therapy. Vectors may be introduced *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors are introduced into stem cells taken from the subject, and the resulting
30 transgenic cells are clonally propagated for autologous transplant back into that same subject. Delivery of the cDNA by transfection, liposome injections, or polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman *et al.* (1997) *Nature Biotechnology* 15:462-466.) Additionally, endogenous gene expression may be inactivated using homologous recombination methods which insert an inactive gene sequence into the coding region or
35 other targeted region of the cDNA. (See, e.g. Thomas *et al.* (1987) *Cell* 51: 503-512.)

WO 02/18575

PCT/US01/26682

Vectors containing the cDNA can be transformed into a cell or tissue to express a missing protein or to replace a nonfunctional protein. Similarly a vector constructed to express the complement of the cDNA can be transformed into a cell to downregulate the protein expression. Complementary or antisense sequences may consist of an oligonucleotide derived from the transcription initiation site; nucleotides between about positions -10 and +10 from the ATG are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, enhancers, repressors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee *et al.* In: Huber and Carr (1994) *Molecular and Immunologic Approaches*, Futura Publishing, Mt. Kisco NY, pp. 163-177.)

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the cleavage of mRNA and decrease the levels of particular mRNAs, such as those comprising the cDNAs of the invention. (See, e.g., Rossi (1994) *Current Biology* 4: 469-471.) Ribozymes may cleave mRNA at specific cleavage sites. Alternatively, ribozymes may cleave mRNAs at locations dictated by flanking regions that form complementary base pairs with the target mRNA. The construction and production of ribozymes is well known in the art and is described in Meyers (*supra*).

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiester linkages within the backbone of the molecule. Alternatively, nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases, may be included.

Further, an antagonist, or an antibody that binds specifically to the protein may be administered to a subject to treat a cell cycle disorder. The antagonist, antibody, or fragment may be used directly to inhibit the activity of the protein or indirectly to deliver a therapeutic agent to cells or tissues which express the protein. The therapeutic agent may be a cytotoxic agent selected from a group including, but not limited to, abrin, ricin, doxorubicin, daunorubicin, taxol, ethidium bromide, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicine, dihydroxy anthracin dione, actinomycin D, diphtheria toxin, *Pseudomonas* exotoxin A and 40, radioisotopes, and glucocorticoid.

Antibodies to the protein may be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies, such as those which inhibit dimer formation, are especially preferred for therapeutic use.

WO 02/18575

PCT/US01/26682

Monoclonal antibodies to the protein may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma, the human B-cell hybridoma, and the EBV-hybridoma techniques. In addition, techniques developed for the production of chimeric antibodies can be used. (See, e.g., Pound (1998) Immunochemical Protocols, Methods Mol Biol Vol. 80). Alternatively, techniques described for the production of single chain antibodies may be employed. Fabs which contain specific binding sites for the protein may also be generated. Various immunoassays may be used to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art.

Yet further, an agonist of the protein may be administered to a subject to treat or prevent a disease associated with decreased expression, longevity or activity of the protein.

An additional aspect of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic applications discussed above. Such pharmaceutical compositions may consist of the protein or antibodies, mimetics, agonists, antagonists, or inhibitors of the protein. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a subject alone or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard

WO 02/18575

PCT/US01/26682

pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating and contrasting the ED₅₀ (the dose therapeutically effective in 50% of the population) and LD₅₀ (the dose lethal to 50% of the population) statistics. Any of the therapeutic compositions described above may be applied to any subject in need of such therapy, including, but not limited to, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

EXAMPLES

It is to be understood that this invention is not limited to the particular devices, machines, materials and methods described. Although particular embodiments are described, equivalent embodiments may be used to practice the invention. The described embodiments are provided to illustrate the invention and are not intended to limit the scope of the invention which is limited only by the appended claims.

I cDNA Library Construction

The LUNGTUT09 cDNA library was constructed from cancerous lung tissue obtained from a 68-year-old Caucasian male during a segmental lung resection following diagnosis of malignant neoplasm of the upper right lobe of the lung. Pathology of the right upper lobe of the lung indicated an invasive grade 3 squamous cell carcinoma forming an infiltrating mass involving the bronchus and the surrounding parenchyma. Patient history includes previous diagnoses of type II diabetes without complications, thyroid disorder, depressive disorder, hyperlipidemia, ulcer of the esophagus, and atherosclerosis. Family history included alcohol use in the mother and father, atherosclerosis in a sibling and a grandparent and malignant brain neoplasm in the mother.

The frozen tissues were homogenized and lysed in TRIZOL reagent (1 g tissue/10 ml; Life Technologies), using a POLYTRON homogenizer (Brinkmann Instruments, Westbury NY). After a brief incubation on ice, chloroform was added (1:5 v/v), and the lysate was centrifuged. The upper chloroform layer was removed to a fresh tube, and the RNA extracted with isopropanol, resuspended in DEPC-treated water, and treated with DNase for 25 min at 37C. The RNA was re-extracted once with acid phenol-chloroform, pH 4.7, and precipitated using 0.3M sodium acetate and 2.5 volumes ethanol. 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 SUPERSCRIPRT plasmid system (Life Technologies). The cDNAs were fractionated on a SEPHAROSE CL4B column (Amersham Pharmacia Biotech (APB), Piscataway NJ), and those cDNAs exceeding 400 bp were ligated into pINCY plasmid (Incyte Genomics, Palo Alto CA). The plasmid was subsequently transformed into DH5 α competent cells (Life Technologies).

II Isolation and Sequencing of cDNA Clones

Plasmid DNA was released from the cells and purified using the REAL PREP 96 plasmid kit

WO 02/18575

PCT/US01/26682

(Qiagen). The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile TERRIFIC BROTH (BD Biosciences, San Jose CA) with carbenicillin at 25 mg/l and glycerol at 0.4%; 2) the cultures were incubated for 19 hours after the wells were inoculated and then lysed with 0.3 ml of lysis buffer; 3) following isopropanol precipitation, the DNA pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were transferred to a 96-well block for storage at 4C.

The cDNAs were prepared using a MICROLAB 2200 system (Hamilton, Reno NV) in combination with DNA ENGINE thermal cyclers (MJ Research, Watertown MA). The cDNAs were sequenced by the method of Sanger and Coulson (1975; J Mol Biol 94:441f) using ABI PRISM 377 DNA sequencing systems (ABI). Most of the sequences were sequenced using standard ABI protocols and kits (ABI) at solution volumes of 0.25x - 1.0x. In the alternative, some of the sequences were sequenced using solutions and dyes from APB.

III Selection, Assembly, and Characterization of Sequences

The sequences used for co-expression analysis were assembled from EST sequences, 5' and 3' long read sequences, and full length coding sequences. Selected assembled sequences were expressed in at least three cDNA libraries.

The assembly process is described as follows. EST sequence chromatograms were processed and verified. Quality scores were obtained using PHRED (Ewing *et al.* (1998) Genome Res 8:175-185; Ewing and Green (1998) Genome Res 8:186-194), and edited sequences were loaded into a relational database management system (RDBMS). The sequences were clustered using BLAST with a product score of 50. All clusters of two or more sequences created a bin which represents one transcribed gene.

Assembly of the component sequences within each bin was performed using a modification of Phrap, a publicly available program for assembling DNA fragments (Green, P. University of Washington, Seattle WA). Bins that showed 82% identity from a local pair-wise alignment between any of the consensus sequences were merged.

Bins were annotated by screening the consensus sequence in each bin against public databases, such as GBpri and GenPept from NCBI. The annotation process involved a FASTn screen against the GBpri database in GenBank. Those hits with a percent identity of greater than or equal to 75% and an alignment length of greater than or equal to 100 base pairs were recorded as homolog hits. The residual unannotated sequences were screened by FASTx against GenPept. Those hits with an E value of less than or equal to 10^{-6} were recorded as homolog hits.

Sequences were then reclustered using BLASTn and Cross-Match, a program for rapid amino acid and nucleic acid sequence comparison and database search (Green, *supra*), sequentially. Any BLAST alignment between a sequence and a consensus sequence with a score greater than 150 was

WO 02/18575

PCT/US01/26682

realigned using cross-match. The sequence was added to the bin whose consensus sequence gave the highest Smith-Waterman score (Smith *et al.* (1992) *Protein Engineering* 5:35-51) amongst local alignments with at least 82% identity. Non-matching sequences were moved into new bins, and assembly processes were repeated.

5 IV Description of the Known Cell Cycle Genes

Genes known to be involved in disease processes involving the cell cycle were selected to identify cDNAs. The known genes and a brief description of their functions are found below.

Gene ID	Name	Description
10 995529	CDC2	CDC2, cell division cycle protein 2 (or cyclin B1) is a mitotic kinase which triggers entry into mitosis. CDC2 binds chromatin prior to S-phase, and is displaced during DNA replication. (Krude <i>et al.</i> (1996) <i>J Cell Sci</i> 109:309-318; De Souza <i>et al.</i> (2000) <i>Exp Cell Res</i> 257:11-21)
15 336106	CDC7	CDC7, cell division cycle protein 7 is a kinase conserved in eukaryotes from yeast to humans. It is essential for initiation of DNA replication and entry into S-phase. (Donaldson <i>et al.</i> (1998) <i>Genes Dev</i> 12:491-501; Jiang <i>et al.</i> (1999) <i>Embo J</i> 18: 5703-5713; and Masai <i>et al.</i> (1999) <i>Front Biosci</i> 4: D834-840)
20 256671	CDC23	CDC23, cell division cycle protein 23, is a component of the anaphase-promoting complex that regulates mitosis by catalyzing the formation of cyclin B-ubiquitin conjugates, targeting cyclin B for degradation. (Prinz (1998) <i>Curr Biol</i> 8:750-760; Zhao <i>et al.</i> (1998) <i>Genomics</i> 53:184-90; and Hershko (1999) <i>Philos Trans R Soc Lond B Biol Sci</i> 354:1571-1576)
25 286623	Cyclin B	Cyclin B is a subunit of cyclin-dependent kinase (cdk) 1. Degradation of cyclin B by the anaphase-promoting complex is required for inactivation of the kinase and exit from mitosis. CDKs are regulators of cell cycle progression and alterations and deregulation of CDK activity are characteristic of neoplasia. CDK inhibitors and modulators alter cell cycle and induce apoptosis and tumor regression. (Hajduch <i>et al.</i> (1999) <i>Adv Exp Med Biol</i> 457:341-53; Hershko, <i>supra</i> ; and Sausville (1999) <i>Pharmacol Ther</i> 82:285-92)
30 392739	hBub1	hBub1, a mitotic checkpoint kinase, is a kinetochore protein that monitors chromosome attachment to the spindle in mitotic cells and controls exit from mitosis and chromosome segregation. The mitotic checkpoint ensures proper chromosome segregation by delaying anaphase until chromosomes are aligned on the spindle. Following spindle damage, cells exit mitosis and undergo apoptosis. hBub1 is required for the checkpoint response to spindle damage; mutations in hBub1 disrupt the mitotic checkpoint allowing cells to escape apoptosis and continue cell cycle progression, despite spindle damage, potentially leading to aneuploidy and contributing to neoplasia. (Taylor and McKeon (1997) <i>Cell</i> 89:727-735; Cahill (1998) <i>Nature</i> 392:300-303; Ouyang <i>et al.</i> (1998) <i>Cell Growth Differ</i> 9:877-885; Imai <i>et al.</i> (1999) <i>Jpn J Cancer Res</i> 90:837-840; Seeley <i>et al.</i> (1999) <i>Biochem Biophys Res Commun</i> 257:589-595; and Myrie <i>et al.</i> (2000) <i>Cancer Lett</i> 152:193-99.
35 337334	hKSP	hKSP, kinesin-like spindle protein (HsEg5), is a spindle-associated

WO 02/18575	PCT/US01/26682
	protein found with centrosomal microtubules during prophase and prometaphase centrosome separation, and associated with post-mitotic centrosome movement. (Whitehead <i>et al.</i> (1996) <i>Cell Motil Cytoskeleton</i> 35:298-308)
5 201204	hp55cdc
	hp55cdc is a kinetochore and spindle microtubule-associated protein that mediates association of the spindle checkpoint protein Mad2 with the cyclosome/anaphase promoting complex and is essential for cell division. Over expression of p55cdc induces apoptosis. hp55cdc is also associated with the mitotic spindle protein kinase Aik. (Weinstein <i>et al.</i> (1994) <i>Mol Cell Biol</i> 14:3350-3363; Kao <i>et al.</i> (1996) <i>Oncogene</i> 13:1221-1229; Kallio <i>et al.</i> (1998) <i>J Cell Biol</i> 141:1393-1406; Kramer <i>et al.</i> (1998) <i>Curr Biol</i> 8:1207-1210; Farruggio <i>et al.</i> (1999) <i>Proc Natl Acad Sci</i> 96:7306-7311; and Saffery <i>et al.</i> (2000) <i>Hum Mol Genet</i> 9:175-85)
10	
15 331025	MCAK
	MCAK, mitotic centromere-associated kinesin, is a microtubule motor protein recruited to the centromere at prophase that participates in anaphase chromosome segregation. (Kim <i>et al.</i> (1997) <i>Biochim Biophys Acta</i> 1359:181-186; Maney <i>et al.</i> (2000) <i>Int Rev Cytol</i> 194:67-131; Maney <i>et al.</i> (1998) <i>J Cell Biol</i> 142:787-801; Wordeman <i>et al.</i> (1999) <i>Cell Biol Int</i> 23:275-86; and Saffery, <i>supra</i>)
20	
26662	mitosin
	Mitosin (CENP-F kinetochore protein) is a nuclear protein that associates with centromeres and spindle poles during M phase. Overexpression of N-terminally truncated mitosin blocks cell cycle progression. Mitosin is correlated with clinical outcome in node-negative breast cancer. (Clark <i>et al.</i> (1997) <i>Cancer Res</i> 57:5505-08; Zhu (1999) <i>Mol Cell Biol</i> 19:1016-1024; and Zhu <i>et al.</i> (1997) <i>J Cell Biochem</i> 66:441-449)
25	
412661	mki67a
	mki67a (MIB-1) is a definitive cell proliferation marker. It is widely used in pathology to measure the growth fraction of cells in human tumors. (Schluter <i>et al.</i> (1993) <i>J Cell Biol</i> 123:513-522; Duchrow <i>et al.</i> (1995) <i>Arch Immunol Ther Exp</i> 43:117-121; Dalquen <i>et al.</i> (1997) <i>Acta Cytol</i> 41:229-237; and Scholzen and Gerdes (2000) <i>J Cell Physiol</i> 182:311-322)
30	
319885	MKLP-1
	MKLP1, mitotic kinesin-like protein 1, is a spindle-associated protein required for mitotic progression. (Nislow <i>et al.</i> (1992) <i>Nature</i> 359:543-7; Sharp <i>et al.</i> (1997) <i>J Cell Biol</i> 138:833-843; Kobayashi <i>et al.</i> (1998) <i>J Cell Biol</i> 143:1961-70)
35	
977509	myb
	B-myb is a member of the myb family of cell-cycle regulated transcription factors, expressed in G1 and S phase. Activity of b-myb is stimulated by cyclin A/cdk2-dependent phosphorylation. (Robinson <i>et al.</i> (1996) <i>Oncogene</i> 12:1855-64; Saville and Watson (1998) <i>Adv Cancer Res</i> 72:109-40; Saville and Watson (1998) <i>Oncogene</i> 17:2679-2689; and Horstmann <i>et al.</i> (2000) <i>Oncogene</i> 19:298-306)
40	
45 336560	NLK1
	NLK1, NIMA-like protein kinase 1, is a human mitotic kinase, similar to the NIMA cell-cycle regulatory protein kinase in <i>Aspergillus</i> that is essential for entry into and progression through mitosis. (Lu and Hunter (1995) <i>Cell</i> 81:413-424; Lu and Hunter (1995) <i>Prog Cell Cycle Res</i> 1:187-205; and Shen <i>et al.</i> (1997) <i>Proc Natl Acad Sci</i> 94:13618-13623)
50	
347876	P1-CDC21
	P1-CDC21 is a member of the family of minichromosome maintenance proteins essential for DNA replication. (Hu <i>et al.</i> (1993) <i>Nucleic Acids Res</i> 21:5289-5293; Ishimi <i>et al.</i> (1996) <i>J Biol</i>

WO 02/18575		PCT/US01/26682
411205	PRC1	Chem 271:24115-24122) PRC1, protein regulating cytokinesis 1, is a human mitotic-spindle associated CDK substrate protein required for cytokinesis. (Jiang <i>et al.</i> (1998) <i>Mol Cell</i> 2:877-885)
5 348211	Aik2	The protein kinase Aik2 / Aurora2 is localized to the mitotic spindle poles, involved in regulating chromosome segregation and maintaining genomic stability, and associated with p53cdc/cdc20. (Kimura <i>et al.</i> (1999) <i>J Biol Chem</i> 274:7334-40; Kimura <i>et al.</i> (1998) <i>Cytogenet Cell Genet</i> 82:147-52; and Farruggio, <i>supra</i>)
10 251651	survivin	Survivin is an apoptosis inhibitor expressed in the G2/M phase of the cell cycle. At the beginning of mitosis it associates with microtubules of the mitotic spindle. It inhibits apoptosis allowing cancer cells to survive. (Li <i>et al.</i> (1998) <i>Nature</i> 396:580-584; Verdecia <i>et al.</i> (2000) <i>Nat Struct Biol</i> 7:602-608)
15 232888	topo II	Topoisomerase II is required for chromosome condensation and segregation during DNA replication. Its expression is cell cycle dependent; both protein level and catalytic activity peaks in G2/M. As part of the regulatory checkpoint at the entry and progression of mitosis, it regulates apoptosis. Topoisomerase poisons induce carcinogenic chromosomal alterations. (Holm <i>et al.</i> (1989) <i>Mol Biol Biol</i> 9:159-168; Kaufmann (1998) <i>Proc Soc Exp Biol Med</i> 217:327-334; Summer (1995) <i>Exp Cell Res</i> 217:440-447; Anderson and Roberge (1996) <i>Cell Growth Differ</i> 7:83-90; Larsen <i>et al.</i> (1996) <i>Prog Cell Cycle Res</i> 2:229-239; and Cimini <i>et al.</i> (1997) <i>Cytogenet Cell Genet</i> 76:61-67)
20 235191	UbcH10	Cyclin-selective ubiquitin carrier protein (UbcH10/E2-C) catalyzes the ubiquitin-mediated proteolysis of mitotic cyclins and is required for cells to complete mitosis and enter anaphase of the next cell cycle. Mutant UbcH10 inhibits the destruction of cyclins, arrests cells in M phase, and inhibits the onset of anaphase. (Towusley <i>et al.</i> (1997) <i>Proc Natl Acad Sci</i> 94:2362-2367; Bastians <i>et al.</i> (1999) <i>Mol Biol Cell</i> 10:3927-3941)

35 V Co-expression Analyses of Known Cell Cycle Genes

Using the LIFSEQ GOLD database (Dec99, Incyte Genomics), we have identified ten cDNAs that show strong association with known cell cycle genes. Initially, degree of association was measured by probability values using a cutoff p-value less than 0.00001. This was followed by annotation and literature searches to insure that the genes that passed the probability test had strong association with known cell cycle genes. The process was reiterated so that an initial selection of 37,071 genes were reduced to the final ten cDNAs claimed herein. The entries in the table below are the negative log of the p-value (-log p) for the co-expression of the two genes. The cDNAs are identified by their LIFSEQ GOLD ID numbers, and the known genes, by their abbreviations as shown above and the number assigned in column 1 which is also used in row 1. The single highest p-values between each of the known genes have been marked in bold. The single highest p-values between at least one known gene and each cDNA is summarized in THE INVENTION section.

WO 02/18575

PCT/US01/26682

Name/Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1 CDC2	NA																		
2 CDC23	13	NA																	
3 CDC7	3.4	5.5	NA																
5 4 Cyclin B	12	6.6	10	NA															
5 hBub1	7.8	7.7	0	5.2	NA														
6 hKSP	5.8	6	5.2	10	4.9	NA													
7 hp55cdc	5.7	4.8	5.8	12	5.5	7.2	NA												
8 MCAK	4.8	6.9	9.2	14	6.1	8	11	NA											
10 9 mitotin	9.8	4.6	6.6	14	12	7	13	12	NA										
10 mki67a	12	5.5	8.3	6.7	6.5	5.8	7.8	11	13	NA									
11 MKLP-1	6.9	5.1	4.5	3.8	6.9	5.3	8.8	3.9	7.1	7.2	NA								
12 myb	0	5.5	7.9	13	5.9	5.7	19	18	13	20	6.8	NA							
13 NLK1	0	4.3	0	0	10	3.5	4.2	4.8	8.3	3.9	3.9	0	NA						
15 14 P1-CDC21	11	7.4	7	10	6.4	5.6	14	9.2	12	21	5.5	10	5.6	NA					
15 PRC1	7.2	5.4	5.9	15	9.7	12	16	12	13	13	4.6	8.5	7.4	15	NA				
16 prkAik2	6.6	8.4	3.2	11	5	6.8	9.3	6.7	7.6	11	5	8.1	3.8	7.2	10	NA			
17 survivin	11	6.5	9.2	11	9.3	9.2	18	11	18	9	10	11	6.1	11	7.9	9.8	NA		
18 topo II	23	11	13	17	11	15	19	14	24	15	6.1	18	8.2	19	18	11	23	NA	
20 19 UbcH10	6.1	5.8	4	12	7.5	12	15	8.7	12	12	5.6	16	6.2	9.4	8	15	16	11	NA
40371	7.9	3.9	5.8	4.6	9.4	6.6	5.5	7.6	7.6	8.8	7.5	7	9.1	8.5	8.7	6.4	5.8	16	10
200394	7.1	5.2	4.9	7.6	6.5	7.5	8.9	6.8	4.5	7.3	9	4.3	4.2	9.5	12	6.6	5.2	11	7.1
201989	5.9	12	4.5	10	4.5	5.4	8.3	9.5	6.7	9.6	3.2	9.7	0	11	8.5	11	8	11	11
211475	9.2	6.6	4.8	7.9	4.5	5.9	8.2	6	5.5	6.7	4.3	4.9	0	9.8	10	7.6	5.4	10	6.2
25 225657	4.7	5	5	13	4.5	4.5	7.9	9.8	11	8.3	3.7	7.2	3.7	6.4	5.5	11	8.4	6.7	13
350770	6.5	6.5	9.9	7.2	5.8	5.9	12	11	9.1	12	0	11	3.7	15	16	8	6.5	14	11
407614	6.6	0	5.5	9.5	0	3.6	7.9	7.6	4.9	4.8	3.8	6.5	0	3.5	5.9	4.6	8.2	9.4	5.6
475113	9	10	4.2	6.7	7.6	8.6	9.1	5.1	9.3	9.4	5.7	10	4.5	9	10	8.4	9	13	12
898622	0	0	0	5.2	5.6	7.9	3	0	4.9	0	0	0	3.3	5.2	4.2	4	6.6	9	3.2
30 978267	0	3.4	10	7.8	3.8	5	17	9.5	8	4.1	4.2	0	4.2	7.8	8.2	5.6	9.4	15	5.6

VI Homology Searching of cDNA Clones and Their Deduced Proteins

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

As detailed in Karlin (*supra*), BLAST matches between a query sequence and a database sequence were evaluated statistically and only reported when they satisfied the threshold of 10^{-25} for nucleotides and 10^{-14} for peptides. Homology was also evaluated by product score calculated as follows: the % nucleotide or amino acid identity [between the query and reference sequences] in BLAST is multiplied by the % maximum possible BLAST score [based on the lengths of query and

WO 02/18575

PCT/US01/26682

reference sequences] and then divided by 100. In comparison with hybridization procedures used in the laboratory, the electronic stringency for an exact match was set at 70, and the conservative lower limit for an exact match was set at approximately 40 (with 1-2% error due to uncalled bases).

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

15 The cDNAs of this application were compared with assembled consensus sequences or templates found in the LIFESEQ GOLD database. Component sequences from cDNA, extension, full length, and shotgun sequencing projects were subjected to PHRED analysis and assigned a quality score. All sequences with an acceptable quality score were subjected to various pre-processing and editing pathways to remove low quality 3' ends, vector and linker sequences, polyA tails, Alu repeats, 20 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 25 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 30 based on the number and orientation of its component sequences.

Bins were compared to one another and those having local similarity of at least 82% were combined and reassembled. Bins having templates with less than 95% local identity were split. Templates were subjected to analysis by STITCHER/EXON MAPPER algorithms that analyze the probabilities of the presence of splice variants, alternatively spliced exons, splice junctions, 35 differential expression of alternative spliced genes across tissue types or disease states, and the like.

WO 02/18575

PCT/US01/26682

Assembly procedures were repeated periodically, and templates were annotated using BLAST against GenBank databases such as GBpri. An exact match was defined as having from 95% local identity over 200 base pairs through 100% local identity over 100 base pairs and a homolog match as having an E-value (or probability score) of $\leq 1 \times 10^{-8}$. The templates were also subjected to frameshift FASTx against GENPEPT, and homolog match was defined as having an E-value of $\leq 1 \times 10^{-8}$. Template analysis and assembly was described in USSN 09/276,534, filed March 25, 1999.

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

The cDNA was further analyzed using MACDNASIS PRO software (Hitachi Software Engineering), and LASERGENE software (DNASTAR) and queried against public databases such as the GenBank rodent, mammalian, vertebrate, prokaryote, and eukaryote databases, SwissProt, BLOCKS, PRINTS, PFAM, and Prosite.

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

VIII Hybridization Technologies and Analyses

Immobilization of cDNAs on a Substrate

The cDNAs are applied to a substrate by one of the following methods. A mixture of cDNAs is fractionated by gel electrophoresis and transferred to a nylon membrane by capillary transfer. Alternatively, the cDNAs are individually ligated to a vector and inserted into bacterial host cells to form a library. The cDNAs are then arranged on a substrate by one of the following methods. In the first method, bacterial cells containing individual clones are robotically picked and arranged on a nylon membrane. The membrane is placed on LB agar containing selective agent (carbenicillin, kanamycin, ampicillin, or chloramphenicol depending on the vector used) and incubated at 37°C for

WO 02/18575

PCT/US01/26682

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).

5 In the second method, cDNAs are amplified from bacterial vectors by thirty cycles of PCR using primers complementary to vector sequences flanking the insert. PCR amplification increases a starting concentration of 1-2 ng nucleic acid to a final quantity greater than 5 μ g. Amplified nucleic acids from about 400 bp to about 5000 bp in length are purified using SEPHACRYL-400 beads (APB). Purified nucleic acids are arranged on a nylon membrane manually or using a dot/slot
10 blotting manifold and suction device and are immobilized by denaturation, neutralization, and UV irradiation as described above. Purified nucleic acids are robotically arranged and immobilized on polymer-coated glass slides using the procedure described in USPN 5,807,522. Polymer-coated slides are prepared by cleaning glass microscope slides (Corning, Acton MA) by ultrasound in 0.1% SDS and acetone, etching in 4% hydrofluoric acid (VWR Scientific Products, West Chester PA),
15 coating with 0.05% aminopropyl silane (Sigma-Aldrich) in 95% ethanol, and curing in a 110C oven. The slides are washed extensively with distilled water between and after treatments. The nucleic acids are arranged on the slide and then immobilized by exposing the array to UV irradiation using a STRATALINKER UV-crosslinker (Stratagene). Arrays are then washed at room temperature in 0.2% SDS and rinsed three times in distilled water. Non-specific binding sites are blocked by
20 incubation of arrays in 0.2% casein in phosphate buffered saline (PBS; Tropix, Bedford MA) for 30 min at 60C; then the arrays are washed in 0.2% SDS and rinsed in distilled water as before.

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
25 prepared by diluting the cDNAs to a concentration of 40-50 ng in 45 μ l TE buffer, denaturing by heating to 100C for five min, and briefly centrifuging. The denatured cDNA is then added to a REDIPRIME tube (APB), gently mixed until blue color is evenly distributed, and briefly centrifuged. Five μ l of [³²P]dCTP is added to the tube, and the contents are incubated at 37C for 10 min. The labeling reaction is stopped by adding 5 μ l of 0.2M EDTA, and probe is purified from unincorporated
30 nucleotides using a PROBEQUANT G-50 microcolumn (APB). The purified probe is heated to 100C for five min, snap cooled for two min on ice, and used in membrane-based hybridizations as described below.

Probe Preparation for Polymer Coated Slide Hybridization

Hybridization probes derived from mRNA isolated from samples are employed for screening
35 cDNAs of the Sequence Listing in array-based hybridizations. Probe is prepared using the

WO 02/18575

PCT/US01/26682

GEMbright kit (Incyte Genomics) by diluting mRNA to a concentration of 200 ng in 9 μ l TE buffer and adding 5 μ l 5x buffer, 1 μ l 0.1 M DTT, 3 μ l Cy3 or Cy5 labeling mix, 1 μ l RNase inhibitor, 1 μ l reverse transcriptase, and 5 μ l 1x yeast control mRNAs. Yeast control mRNAs are synthesized by in vitro transcription from noncoding yeast genomic DNA (W. Lei, unpublished). As quantitative controls, one set of control mRNAs at 0.002 ng, 0.02 ng, 0.2 ng, and 2 ng are diluted into reverse transcription reaction mixture at ratios of 1:100,000, 1:10,000, 1:1000, and 1:100 (w/w) to sample mRNA respectively. To examine mRNA differential expression patterns, a second set of control mRNAs are diluted into reverse transcription reaction mixture at ratios of 1:3, 3:1, 1:10, 10:1, 1:25, and 25:1 (w/w). The reaction mixture is mixed and incubated at 37C for two hr. The reaction mixture is then incubated for 20 min at 85C, and probes are purified using two successive CHROMA SPIN+TE 30 columns (Clontech, Palo Alto CA). Purified probe is ethanol precipitated by diluting probe to 90 μ l in DEPC-treated water, adding 2 μ l 1mg/ml glycogen, 60 μ l 5 M sodium acetate, and 300 μ l 100% ethanol. The probe is centrifuged for 20 min at 20,800xg, and the pellet is resuspended in 12 μ l resuspension buffer, heated to 65C for five min, and mixed thoroughly. The probe is heated and mixed as before and then stored on ice. Probe is used in high density array-based hybridizations as described below.

Membrane-based Hybridization

Membranes are pre-hybridized in hybridization solution containing 1% Sarkosyl and 1x high phosphate buffer (0.5 M NaCl, 0.1 M Na₂HPO₄, 5 mM EDTA, pH 7) at 55C for two hr. The probe, diluted in 15 ml fresh hybridization solution, is then added to the membrane. The membrane is hybridized with the probe at 55C for 16 hr. Following hybridization, the membrane is washed for 15 min at 25C in 1mM Tris (pH 8.0), 1% Sarkosyl, and four times for 15 min each at 25C in 1mM Tris (pH 8.0). To detect hybridization complexes, XOMAT-AR film (Eastman Kodak, Rochester NY) is exposed to the membrane overnight at -70C, developed, and examined visually.

Polymer Coated Slide-based Hybridization

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

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

WO 02/18575

PCT/US01/26682

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

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

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

IX Complementary Molecules

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

WO 02/18575

PCT/US01/26682

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

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

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

X Protein Expression

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

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

XI Production of Antibodies

Tumor antigen is purified using polyacrylamide gel electrophoresis and used to immunize mice or rabbits. Antibodies are produced using the protocols below. Alternatively, the amino acid

WO 02/18575

PCT/US01/26682

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

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

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

XII Purification of Naturally Occurring Protein Using Specific Antibodies

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

XIII Screening Molecules for Specific Binding with the cDNA or Protein

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

XIV Two-Hybrid Screen

A yeast two-hybrid system, MATCHMAKER LexA Two-Hybrid system (Clontech Laboratories, Palo Alto CA), is used to screen for peptides that bind the 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

WO 02/18575

PCT/US01/26682

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

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

XV Transcript Imaging

A transcript image was performed using the LIFESEQ GOLD database (Jun01release, Incyte Genomics). This process allowed assessment of the relative abundance of the expressed cDNAs in more than 1400 cDNA libraries. Criteria for transcript imaging can be selected from category, number of cDNAs per library, library description, disease indication, clinical relevance of sample, and the like.

All sequences and cDNA libraries in the LIFESEQ database have been categorized by system, organ/tissue and cell type. For each category, the number of libraries in which the sequence was expressed were counted and shown over the total number of libraries in that category. In some transcript images, all normalized or subtracted libraries, which have high copy number sequences 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 disregarded or removed where clinical relevance is emphasized. Conversely, fetal tissue may be emphasized wherever elucidation of inherited disorders or differentiation of particular cells or organs from stem

WO 02/18575

PCT/US01/26682

cells (such as nerves, heart or kidney) would be furthered by removing clinical samples from the analysis.

The transcript images for SEQ ID NOs:1, 5, and 10 are shown below. The first column shows library name; the second column, the number of cDNAs sequenced in that library; the third column, the description of the library; the fourth column, absolute abundance of the transcript in the library; and the fifth column, percentage abundance of the transcript in the library.

Category: All (SEQ ID NO:1)

Library*	cDNAs	Description of Prostate Tissue	Abundance	% Abund
CONDUTP01	1286	peritoneum, neuroendocrine CA, 66F	2	0.1555
PENHTUP02	1846	penis squamous cell CA, 64M, 5RF	1	0.0542
LUNGTPU09	3969	lung squamous cell CA, 68M	2	0.0504
OVARTUP02	2932	ovary papillary serous CA, 64F, VM/VM	1	0.0341
SPLNTUP02	3077	spleen Hodgkin's, 45M	1	0.0325
COLLTPU02	6656	ileocecum, Burkitt lymphoma, 29F	2	0.0300

15 *Cell line, fetal, pooled, subtracted and normalized libraries were not used in this analysis.

Differential expression of SEQ ID NO:1 in neuroendocrine carcinoma of the peritoneum is 3-fold greater by percent abundance than expression in any other tissue of the digestive tract. No expression was found in cytologically normal tissue. When used in a cell or tissue specific diagnostic procedure and compared to established standards, SEQ ID NO:1 is diagnostic for cancer, specifically neuroendocrine carcinoma, of the peritoneum.

Category: Exocrine (Breast)

Library*	cDNAs	Description of Bladder Tissue	Abundance	% Abund
BRSTUPF01	1146	breast tumor line T-47D, ductal CA, 54F	1	0.0873
BRSTUP16	3724	breast ductal CA, 43F, m/BRSTTMT01	2	0.0537
BRSTTUP08	3928	breast tumor, adenoCA, 45F, m/BRSTNOT09	2	0.0509
BRSTUP01	3130	breast tumor line T47D, 54F	1	0.0319
BRSTNOT03	6777	m/BRSTTUP02 ductal adenoCA, 54F	1	0.0148
BRSTUP13	7631	breast adenoCA, 46F, m/BRSTNOT33	1	0.0131
BRSTUP03	10092	breast lobular CA, 58F, m/BRSTNOT05	1	0.0099

*No libraries were excluded from this analysis

SEQ ID NO:5 is diagnostic of breast cancer as shown by its expression in breast tumor line T-47D and in these matched sets of cancerous and normal breast tissues. Expression was not found in cytological normal breast tissue removed from subjects during breast reduction surgery or any other breast library. When used with breast tissue, SEQ ID NO:1 is diagnostic for breast cancer.

Category: Digestive Tract (Colon)

Library	cDNAs	Description of Lung Tissue	Abundance	% Abund
COLNTUP12	2312	colon adenoCA, M/F, pool, 3' CGAP	1	0.0433
COLNTUP15	12065	colon adenoCA, pool, NORM, 3' CGAP	5	0.0414
COLNTUP03	2462	colon adenoCA, M/F, pool, NORM	1	0.0406
COLNTUP17	7421	colon adenoCA, 3', CGAP	2	0.0270
COLLTPU02	6656	Burkitt lymphoma, 29F, m/COLANOT03	1	0.0150
COLNTUP16	8499	colon adenoCA, pool, NORM, 3'/5' CGAP	1	0.0118

45 Differential expression of SEQ ID NO:10 was not found in libraries constructed from the tissues of subjects diagnosed with chronic ulcerative colitis (COLADIT05, COLANOT02, COLAUCT01, and COLDDIE01), benign familial polyposis (COLCDIT01, COLDNOT01, and

WO 02/18575

PCT/US01/26682

COLTDIT04), ulcerative colitis (COLNDIP02, COLNNOT23, COLNUCT03, and COLSUCT01), or
in cytologically normal tissue (COLNNON05, COLNNOP01, COLNNOP02, COLNNOT01,
COLNNOT05, COLNNOT07, COLNNOT08, COLNNOT09, COLNNOT11, COLNNOT13,
COLNNOT16, COLNNOT19, and COLNNOT22). When used in a cell or tissue specific diagnostic
5 procedure and compared to established standards, SEQ ID NO:1 is diagnostic for colon cancer.

In assays using established standards and patient samples, the cDNA, an mRNA, a protein or
an antibody specifically binding the protein serves a clinically relevant diagnostic marker for cell
cycle disorders.

10 All patents and publications mentioned in the specification are incorporated by reference
herein. Various modifications and variations of the described method and system of the invention
will be apparent to those skilled in the art without departing from the scope and spirit of the
invention. Although the invention has been described in connection with specific preferred
embodiments, it should be understood that the invention as claimed should not be unduly limited to
15 such specific embodiments. Indeed, various modifications of the described modes for carrying out
the invention that are obvious to those skilled in the field of molecular biology or related fields are
intended to be within the scope of the following claims.

WO 02/18575

PCT/US01/26682

What is claimed is:

1. A composition comprising a plurality of cDNAs having the nucleic acid sequences of SEQ ID NOs:1-10 or the complements thereof.
2. A method for using a composition to detect gene expression in a sample containing nucleic acids, the method comprising:
 - a) hybridizing the composition of claim 1 to the nucleic acids under conditions for formation of one or more hybridization complexes; and
 - b) detecting hybridization complex formation, wherein complex formation indicates gene expression in the sample.
3. The method of claim 2 wherein the cDNAs of the composition are attached to a substrate.
4. The method of claim 7 wherein gene expression is compared to a standard and is indicative of a cell cycle disorder.
5. A method of using a composition to screen a plurality of molecules or compounds, the method comprising:
 - a) combining the composition of claim 1 with a plurality of molecules or compounds under conditions to allow specific binding; and
 - b) detecting specific binding, thereby identifying a molecule or compound that specifically binds a cDNA of the composition.
6. A cDNA comprising a nucleic acid sequence selected from SEQ ID NOs: 1, 2, 4-10 and a complement thereof.
7. A composition comprising the cDNA of claim 6 and a labeling moiety or a pharmaceutical carrier.
8. A method for using a cDNA to detect expression in a sample containing nucleic acids, the method comprising:
 - a) hybridizing the cDNA of claim 6 to the nucleic acids under conditions for formation of a more hybridization complex; and
 - b) detecting complex formation, wherein complex formation indicates expression in the sample.
9. The method of claim 8 wherein the cDNAs of the composition are attached to a substrate.
10. The method of claim 8 wherein expression is compared to a standard and is indicative of a cell cycle disorder.
11. A method of using a cDNA to screen a plurality of molecules or compounds to identify and purify a ligand, the method comprising:
 - a) combining the cDNA of claim 6 with a plurality of molecules or compounds under conditions to allow specific binding; and
 - b) recovering the bound cDNA ;

WO 02/18575

PCT/US01/26682

- c) dissociating the cDNA from the ligand thereby obtaining a purified ligand.
12. The method of claim 11 wherein the plurality of molecules or compounds is selected from DNA molecules, RNA molecules, peptide nucleic acids, transcription factors, enhancers, repressors, mimetics, and proteins.
13. An expression vector comprising a cDNA selected from SEQ ID NOs: 1, 2, and 4-10.
14. A host cell comprising the expression vector of claim 13.
15. A method for using a cDNA to produce a protein, the method comprising:
- culturing the host cell of claim 14 under conditions for protein expression; and
 - recovering the protein from cell culture.
16. A purified protein or a portion thereof produced by the method of claim 15.
17. A composition comprising the protein produced by the method of claim 15 and a labeling moiety or a pharmaceutical carrier.
18. A method for using a protein to screen a plurality of molecules or compounds to identify and purify at least one ligand which specifically binds the protein, the method comprising:
19. The method of claim 18 wherein the plurality of molecules is selected from DNA molecules, RNA molecules, peptide nucleic acids, mimetics, proteins, agonists, antagonists, and antibodies.
20. A method of using a protein to prepare and purify antibodies comprising:
- immunizing an animal with the protein of claim 16 under conditions to elicit an antibody response;
 - isolating animal antibodies;
 - attaching the protein to a substrate;
 - contacting the substrate with isolated antibodies under conditions to allow specific binding to the protein;
 - dissociating the antibodies from the protein, thereby obtaining purified antibodies.

WO 02/18575

PCT/US01/26682

<110> INCYTE GENOMICS, INC.
WALKER, Michael G.
JUNG, Kenneth

<120> GENES EXPRESSED IN THE CELL CYCLE

<130> PB-0015 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/229,253
<151> 2000-08-30

<160> 10

<170> FERM Program

<210> 1
<211> 1970
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 040371.3

```

<400> 1
gggacttcca gtaggaggcg gcatetttga aaagtgatga cggttgacgt ttgctgatt 60
ttgactttgg ttgtagctgc tcccgaact cgcgcgtctc ctgtcagcgg ccggcactgt 120
aggtagagcgc gaggagacgg aggaaggaag cctgagagaca gaagcctctc ccatccaag 180
gcgcggcgca gtcgcgggac gctgggcctg gcggtgtttt cgtcgtgctc agcgggtgga 240
ggagcggcaa gaaaccagag cctgggagat taacagagaaa cttccaagat ggaactttg 300
tctttcccca gatataatgt agctgagatt gtgattcata ttgcgaataa gatcttaaca 360
ggagctgatg gtaaaaacct caccaagat gatctttatc caaatccaaa gcctgaagtc 420
ttgcacatga tctacatgag agccttacia atagtatatg gaattcgact ggaacatttt 480
tacatgatgc cagtgaactc tgaagctatg tatccacatt taatggaagc cttcttacca 540
ttcagcaatt tagttactca tctggaacta tttttgccta tctgcggggt gaatgacttt 600
gagactgctg atattctatg tccaaaagca aaacggacaa gtcggttttt aagtggcatt 660
atcaacttta ttcaacttcag agaagcattc cgtgaaactg atatggaatt tctttggcaa 720
tataaactct ctgcggacaa aatgcaacag ttaaacgcgg cacaccagga ggcattaatg 780
aaactggaga gaactgtatc tgttccagtt gaagagcaag aagagttcaa gcagcttca 840
gatggcaatc agggctaca acaatcaacta aatcaggatt ttctcaaaa aacgatagtg 900
ctcgaagayg gaaattccca aagaagctca aatattcng agaaaaacca gcgtttgaat 960
gaaactaaat tgcgcgtggt ttctttgaaa gaatacaag agagtttgaa aacaaaaatt 1020
gtggattctc ccagacaaga agtggtggag aaatatgaaa tctatggaga ctcagttgac 1080
cttaaaaatg ccagacaaga agtggtggag aaatatgaaa tctatggaga ctcagttgac 1140
tgccctgact ctatgtcagtt ggaagtgcag ttatatcaaa agaaaaatac ggaccttca 1200
gataataggg aaaaattagc cagtatctta aaggagagcc tgaacttgga ggaccaaat 1260
gagagtgatg agtcagaact gaagaantg aagactgaag aaaaatcgtt caaagactg 1320
atgatttgta agaagggaaa acttgccaca gcacaattca aataaataa gaagcattga 1380
gatgtaagc aatacaaaay cacagtaatt gaggattgca ataaagtca agaaaaaaga 1440
ggtagctgct atgacagat aaccacaatt aatcaagaaa tccaaaaaat taaactttga 1500
atcaacaac taagaagatg tctgaaagg gagaacttga agtcccagga aatatttcta 1560
aacttgaaaa ctgcttttga gaataccac gacggatttg aaaaagcagc agaggaactc 1620
tatgttaaga tagatgagaa gncagctgaa ctgaagagga agatgttcaa aatgtcaacc 1680
tgattaacaa aattacatgt ctttttgtaa atggcttgc atcttttaat tttctatata 1740
gaaagaaaag ttgaagcgaa tggaaqtatc agaagtaoca aataatgttg gcttcatcag 1800
ttttttatac ctctcataag tagttaataa gatgaattta atgtaggcct ttatattatt 1860
ataattaaaa taacttgtgc agctattcat gtctctactc tgcaccttgt tgttaaatag 1920
ttgagtaaaa caaaactagt tacctttgaa atatatatat tttttctgt 1970

```

<210> 2

WO 02/18575

PCT/US01/26682

<211> 1570
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 200394.1

<400> 2
cttaaaaagt tgcagaaga agaaggaaa gggaaagaaa agtgttcaga aatccttata 60
tggggaaaga gacattgctt ctaagaagcc cctcctcagt cctattcccg agctgcctga 120
agtccctgag atgacacctt ccatlccgag catccgaaga ctgggttcag gttatctcag 180
ttcaaatggc aaactggaag aagtgaagac tccataaaat ccagtgaaaa gaaaggatct 240
tttgcgtcat gacccagatt tgcataatga tcaaggctat gataaatatg atgtctctga 300
attcgtctct gatataaaaa gttcctcctc gcttggcaat gctacttctg atgaagatcc 360
aaatacaaat ataataaaca ttaataaaaa taaaaatatt ccaaaaagcaa aaaaataagtc 420
agaaagtga aatgaaccaa aagctggaac tgacagctct gtttcttctg cttctataac 480
tgaagaacgt tgggcatcag atagtcctca accctgctct accctgcaag agggctcaaga 540
atttctgctt ggtggtcaaa atgcagaaaa ccttctgctc tctcttaaaa ttccaccaga 600
tttaaacata aagtgtgaaa gaaggatga cttcttagga gctgcaagag gaaaactgca 660
atgcaatcgt ttaatgctca attcacaaaa agactgctat tgttttagag atgtctctaat 720
tgaataatag aaagaactca aaagccagag tgaggatttg ggaagaaac ccatggaag 780
tggcagctgt gtgagttgca gagccagaa agatagaaga cgttccatgt gttatctcga 840
tggtagaagt ttacatttgg aaaaaatgg aaatcacaca ccatcctcca gttgtggcag 900
ctctctagaa attagtttag aaaattctga actgttbaaa gatttctctg atgccattga 960
gcaaaccttt cagaggagaa atagtgaaac caaagtgcga cgtagcaaga ggcctacaga 1020
ggatttagaa aacgaaggtc ttgtatggat ttcacttcca cttcctcca cttcccaaaa 1080
agccaaaaga agaacaatat gtacatttga cagcagtgga ttgaaagta tgtctccat 1140
aaaaaact gtgtcctcca gacaaaacc gcagatggca cctccctctc cagatccaga 1200
aaacagccag ggcctgctg ctggttcttc cgtatgaact ggttaagaga ggaagagctt 1260
ttgtatct acacttgcga atactaaagc cacttcccag tccaaggct accgagagag 1320
atctctctt atgggaaag gagagagctc tctgactgoc ttgaaagga ttgaacata 1380
tgggaaaga aagcagtaat tgacatttcc tgcagagctc gtagcaagag ggaagtaac 1440
catctatgct gaatgatct gtctagtctc cacttctctg tcaacctcag tgtttcaaaa 1500
gttctcaata aataaactca tttgagttga acctactttt atgtagaat aataagttt 1560
cttcatcatt 1570

<210> 3
<211> 1324
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 201989.4

<400> 3
ctgtttgca tccagaggtg gaattggggc ccgtaagtg atttgaataa ttaataaat 60
aagttagagg gctcagcagg cccagaacga gccattttgt cagctgcagc agtcattaac 120
tccgagagg cctctggtcc ctgcccagga agtttcttca ctggaaactg ggaagacagg 180
gtggtttga acttccggag ttgagccagc agctgttctg catccagagg tgaatctggg 240
gcccggcatt cctcctcgt cccgggtcgg ccttgcctcc acacctgca actcctggtt 300
gagatgggct cagccaagag cgtcccagtc acaccagccg ggcctccgcc gcacaacaag 360
catctggctt gagtggcgga cccccgttca cctagtctg gcatcctgag cactcccatc 420
caggtgaga gcttccaca gccaggccta ccagcagggg agcaactgga ggtctttaa 480
catgcccagg actcagatcc ccgctcctct actcttggta ttgacggac acctatgaa 540
accagcagtg gagaccctcc aagcccactg gtgaaacagc tgagtgagat attgaaact 600
gaagactcta aatcaaatct tccccagag cctgttctgc ccccagaggc accttatct 660
tctgaattgg acttgcctct gggtaaccag tlatctgttg aggaacagat gccaccttgg 720
aacccagact agttcccctc caaacagggt ttttccaaag aggaagcaag acagcccaca 780
gaaacccctt tggccagcca gagctccgac aagccctcaa ggaacccctg gactcccaga 840
tcttcagggt ctatgagcaa tagatggaaa caaacagca gcaaggtact agggagatcc 900
cccctcacca tctgagcaga tgacaactcc cctggcacc ctagactacg acagggtaag 960
ggccttcac ccctaagtga aatgttagt gaactaaagg aaggagccat tcttggaaact 1020

WO 02/18575

PCT/US01/26682

```

ggagcacttc tgaaaactgg aggacgagca tgggagcaag gccaggacca tgcaaggaa 1080
aatcagcact ttcccttggg ggaagactag gccctgcatg gccocagcaa tgcagtcacc 1140
cagggcctgg tgsatctgt gtcctctcac cccctctctc ccagggatac tgaggaatgg 1200
cttgttttct tagactcctc ctocagctacc aaactgggac tcaagctttt atctggcttt 1260
ctttgtgtct tgtgtgtttc ttttatatta aaggaagtaa ttttaaatgt tactttaaaa 1320
aggc

```

```

<210> 4
<211> 1857
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 211475.1

```

```

<400> 4
ggagggttcg aattgcaacg gcagctgcog ggcgtatgtg ttggtgctag aggcagctgc 60
agggtctcgc tggggccgcg tggggaccaa ttttgaagag gtacttgcc acgacttatt 120
ttcaactcog accttctctt ccaggcgggt agactctgga ctgagatgg ctttccaat 180
ggaagggatc agtaatttca agacaccagc caaattatca gaaaaaaga aatctgtatt 240
atgttcaact ccaactataa atatcccggc ctctccgttt atgcagaagc ttggtttgg 300
tactggggta aatgtgtacc taatgaaaag atctccaaga ggtttgtctc attctccttg 360
ggctgtaaaa aagattaatc ctatatgtaa tgcattatcg cgaagtgtgt atcaaaagag 420
actaatggat gaagcaaga ttttgaaaag ccttcatcat ccaaacattg ttggttatcg 480
tgcctttact gaagccaatg atggcagctc gtgtcttggc atggaatag gaggtaaaa 540
gtcttcaaat gacttaatag aagaacgata taaagccagc caagatcctt ttccagcagc 600
cataatctta aaagtgtgct tgaatagcc aagaggggta aagtatctgc accaagaaaa 660
gaaatgctt catggagaca taaagctctc aaatgttcta atlaaaggcg attttgaac 720
aatlaaactc tggatgtgag gactctctct accactggat gaaatatacg ctgtgactga 780
ccctgaggtc tgttacattg gcacagagcc atggaaacc aaagaactg tggaggagaa 840
tgggttattc actgacaagg cagacatatt tgcctttggc cttacttttg gggaaatgat 900
gactttatgc attccacaca ttaactcttc aaatgatgat gatgatgaag ataaaacttt 960
tgaatgaaat gatltttgat atgaagcata ctatgcagcg ttggaaacta ggcaccctat 1020
taaatggaa gaactggatg aatcacaaca gaaagtaatt gaactctctc ctgtatgcac 1080
taatgagagc cctaagatgc gtccctctgc tgcacacatt gttgaagctc tggaaacaga 1140
tgtctagtga tcatctcagc tgaagtgtgg ctgacataaa taactgttta ttccaaaaata 1200
ttcaatagat tactatcagt agttattaga ctctaaaatt ggcatttttg aggcacatag 1260
ttctttgta acatatggat aactatctt aatatgaaat atgcttata tggctataag 1320
cacttggaaat tgtactgggt ttctgtttaa gttttgaaa ctagctacat aagtactttg 1380
atactgtca tgetgactca aaacactagc agtaaaacgc tgaactgtt aacttaaat 1440
tgaatgacca ttacttttat taatgactct tcttaaatat tctatatttt aatggactca 1500
ctgacattag cactttgtac agtcaaaat aaagtctaca tttgtttaa acactgaaac 1560
ttttgtgat gtgtttatca aatgataact ggaagctgag tgaatatagc ctcaaaaaga 1620
gtagctcctt ggatacttca gactctggtt acagattgtc ttgactctct ggatctcctc 1680
agatcttgg tttttgcttt aatttattaa atgtatttct catactyagt ttaaaattta 1740
ttaatttga ccttaagcat ttcccagctg tgtaaaaaca ataaaactca aataggatga 1800
taagaataa aggacacttt ggttaccaga aggtgtctca gcattatttt atacttc 1857

```

```

<210> 5
<211> 2447
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 225657.4

```

```

<400> 5
ctccttcttc agcggcggga agctggcggc agcggcgggt gcgggtggctg agcagaggac 60
ccggcggggc gcctcggggg tcaggacaca atgtttgcac gaggactgaa gaggaatgt 120
gttggccacg aggaagactg ggaaggagcc ctggccggct tgaagacagt gtctctatac 180
agcttgagcg gcaagctgct cctggacatg tctctggtag agttgcagct ttgccacatg 240
cttgtggagc caaacctgtg ccgtctcagt ctcaattgcca acacggctcg gaagatccaa 300

```

WO 02/18575

PCT/US01/26682

```

gaggagatga cgcaggatgg gacgtggcgc acagtgccac cccaggtgc agagcggggc 360
ccgctcgacc gcttgggtctc cacggagatc ctgtgcccgtg cagcgtgggg gcaagagggg 420
gcacatcctgt ctccctggctt gggggacggc cacacacagg gtcacgttcc tgacctttgc 480
ccagtcacctc cagcacaggc accaaggcac ctgcagagca gcgcctggga gatggatggc 540
cctcggagaa acagaggaag ctttcacaag tccacttgalc agatatttga aacgctggag 600
actaaaacc cccagctgat ggaagagctg ttctccagac tggacaagcc ctactacgac 660
ctggacaacg tactgacagc catgatgggg ggtgccaggg cgggcccctg cgaagggctc 720
gagggtctgg ctccggccac cccaggccct agctccagct gcaagtcoga cctggggcag 780
ctggaccacg tgatggagat cctggtggag acctgagcag gagccctgag tgcctcacgc 840
cgectctgac gcattgacac gtgagcaact gctcccacgg aggtgtgcgc tgcgccccagc 900
ggcccagcct tgctgcccgt tctgtgtatt ctgagaaatc ccagaacagc ccattaccag 960
tggggctgca gccctaggcc cgtcccactc acctcccccc tgtggaaggg caggcagagg 1020
ctgttctgga aggctctctt tctctgacg tccccacagc cctggggccc tctgtctct 1080
ttgtgtcccc cactgtagag gacggtgagc cgcagctgca tcaacctcct ttaccttta 1140
gataggtgaa tttttacaat tcagttttac atgtttcggg cagtattttg tottaagata 1200
tattttltaa actttttata ccttatctct ttagattttt tcagctattt tottaaaagt 1260
atatttttcc tataaacatc ctttgcctct acctagaaac ttttatagcc taacaatgg 1320
cagttgtgtg gtttcatctt ttttaagttt aaataagggg tttttgtttt gttttgtttt 1380
ttcagtgtag cctcactaca gtctcagta acagtgtaga tggatcatgt tttactctaa 1440
atgtgtgtgt gatactctct cattatgtcc tgcgctgagc tggagacctg gtgaaatca 1500
ggaaccgcac acagccacat ctctctagac ctaagagtaa attatggagg attttatata 1560
tgcctattta tatgtaaatg tcatgaaaga caaaggtcaa atattgtct gttttagat 1620
caagggcacc agttgtctct cagggacctc atagcccctc ggtggtgcct tctcaaggca 1680
gtgttctctg aggtcctcct cagggtcagc ccatgcaact gccctgggtg aggaagtagc 1740
attgtctgct gatgagaaac gccctgcgtg ctctgtttaga ctggtgctga aacaaaagg 1800
taaggctagc ttgaagtcta gaatgaaaga aatctgaatc catgtcatc ataacccctt 1860
gatctgtagc gtcctgggtg ctgcccaggg caggagatga gctgggggtg cctgcagcct 1920
tccactctct ccccgcctca ccccacatgc tccctgttcc tcaatgcttc tctaaactcc 1980
tcacccttca accaaaagg tgtgttttct tttgtgcata tagccattct taatatcag 2040
tgatgtaaac ctaccttat taaaaatta tccagcaaac aaaaaggaa tgggtgttta 2100
gttagacc ccggctgac cctcagcaa ccttctgca ggttagtgc tgtgtatta 2160
tctgtgtgtg ctcttaagg tgggaaaagg aattgcactt ggtctgaaba aatggagcgt 2220
gggttaacttt tatttcccc cccacagggg tgcagagcaa attctttta cattgttcag 2280
gcgccgctg ggggtggggg tgtccacgac ctctgacagc ccccgatgct gaaagttaat 2340
cctcatggac cctagttaa aggtatgta ttttatagga ataatctaa agcactattt 2400
tgtttctgta tagcattttt atcttttaga aacatcattt gttcagc 2447

```

```

<210> 6
<211> 2482
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 350770.3

```

```

<400> 6
gagagtgccc ttcccgggtg gcgcgcgcgc gggggcggcg cgctggagga gctcagagac 60
gagcctagtt atgtctggga ggcgaacgcg gtcgggagga gccctcagc gctccgggcc 120
aagggcccaca tctcctacta agcctctcgg gagggtcccag cggaaatcag gctctgaact 180
cccagacact ctccctgaaa tctggccgaa gacaccaggt cgggctgcag tcagaaaaggc 240
catcgtctta aagaggatcg tggcccactg tgtagaggtc ccagctgtcc aatcacctcg 300
caggagccct aggatltcct ttttcttggg gaagaagaaa gagcccctg gcaggagact 360
tactaaggag gaacttttca agacacacag cgtccctgcc acccccacca gcaactctgt 420
gccgaacctt gaggcagagt ccagctccaa ggaaggagag ctggagcga gagacttggg 480
aatgtctaa agagtcaggc gttctacag cggctggag accctgggg cctcgtcct 540
acctccacc caggcccgcg gtcctgttct ggcttcgagg ggctgtgg ggccagagac 600
ttgtccggag tctcgcaggt ggtgtctcc aaactcacc aggtcccag ggtttgtgca 660
aagccctggg ccccagacat gactctcctt ggaatctccc accaccgga gaaacagaaa 720
cgtaagaaga agaaaatgac agagatctgt aaaaaggagc tggatgagt ggtcgggcc 780
atgaatgoc agtttgaagc tctgagcag tttgatctcc tgggtgaatg agatgcagtg 840
ggggatgca ctggccagac tctccctcct gtcctgtaca tagccacctc cctgtggaga 900
ggacacttag ggtcccctcc cctggctctg ttacctgtgt gbtgtgtgt gctgcgcatg 960
aggactgtct gcctttgagg gcttggggcag cagcggcagc catcttgggt ttaggaaatg 1020

```

WO 02/18575

PCT/US01/26682

```

gggcccgcctg gccccagccac tcactgggtgt cotgetcttg tctctctgtc ctctctatct 1080
ccccaaagta ccatagccag ttccagatg ggccacagac tggggaggag aatcagtgcc 1140
ccagccagaa gttaaagggc tggaggttga ggtgagagcc acctctgcctc ttgttgggag 1200
gggtggctgc ttgaaatag gccccagggcc tctgccagcc tcggcctctc cctcctgagt 1260
tgcttctgtg ttgtggcttt ctctctgaac ccacctgtgt aaagaggttt tcagttccgt 1320
gggtttcccc ttgtattctg taaatagtcc cagagagaaat tctgggctg agggcaatcc 1380
tgctctggag gaagagctg gacatcagc ctgtggagtc tgggtttga aggatgtagg 1440
gagccttagt tgggtctcag accabaagtg tctctccacc agaagotgtg tttcttagtt 1500
ctgtctgctg gttgagatg ttgtaaatg ccaggttgat agggcgctgg ctgcttggag 1560
caaaaggtgc atttcaggtg gtggccacca ggtgctgtga gtttctgtgg ctcactgctc 1620
ctggcctggt cccttgccca gggcccacgc tggagtctta ccactctgct gcagggttgg 1680
aaggtggccc ctcttgccac ccataccact ttcttcaaaa ataagttaaa ccgagtctac 1740
ttggccctag aagagaaagt tgaagagtc ccagacctact agcattttgc aactatgctt 1800
gtaagtcctc cggaaagttt cctcgcgtac cagacacggg cgggggtgta tagcaatttt 1860
agtttttggc ctccctatcc tctcacatga gaacactggc tggatgcac tcatgatctc 1920
tggagaattt cccctctttt ctctctcttc catcgtgtgg attcaatagt gtggatttga 1980
aggctgcctc gcccccgact ctctgcgcgc acccctggcc attgtacctt ttgatgttta 2040
gaagtctgtg gaagttagag ctgaggtgtg cagagggact ggtgataaac agagaatgcc 2100
agggaaagtg agtctgggtt caggttactt ggatgaaacg gtccagccca ggcgggacct 2160
aaataaaccc tctgccaggt ctgggagctc caggccatct ctctaaacgt ctgtggtttg 2220
tcagacctgc aagcaagccc ctctctgggg agccttaggt gtcttgagc tgaaccgca 2280
tgaagaactc ttgtctcac tggctgatgc agcagaactc ttgggaaatg tcttagtctc 2340
gcagaatcag ggtcaccag atgatccaga gttgagatca tcaattgaaa gttctctgtt 2400
cctggggaac taaatttaag gaaaaaatgy gattttgttt tagagttgga aaaaaaacct 2460
gattaaagag tttctgctgt tt 2482

```

```

<210> 7
<211> 2405
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 407614.1

```

```

<400> 7
aagggacttc tcccgaccoc cactctgtcc caggacatag ggcagggggc ctactgctc 60
tgtttggttc caccctgttc ctacctctgc aggcctcttt gctctccctc cttgctccag 120
gaaaccggct ggcaacctgt gctccaagtg actgtcttga acagagcggg ctcttctcatg 180
gctgctgtgt tctgagtttt gaactgctcc tccctggcct gcgtgactga atccagctt 240
ttgtccctgt ctbgcagggg ctgaggtgtc agggagggac ttctggccca ccttgctctc 300
agccctggag tgggcagaga gtattgtggg gaggcatggc cagtgggact agtgttccct 360
ccatctggcc acagctcttg ggagatgggg tgggcagggg tggctcctggc tggcatbgcc 420
tgaaccggcc agtgatgaag tggggagctt gcccttgaca ggttggggct ggtctgggcc 480
ttaatgtgaa aagcagctag ccggcagctg ggttagagcg agccaccag ccctaaaagg 540
ctgctctcat ggccatctag cccagttcca gggcagctc catagccca agccagcgt 600
gggtggggcg ggggtgttcc cacagctggg ttccactga agagcctccg tgctctggag 660
.caggagagcg aggctatggc tgccaccctc cctcctgect ggttcccagt gagaactgac 720
ctgagtcctc ttccaaacc agaccacct cctgccccag gccactgaa gcatgttcca 780
tttctaaaaa gcccaagatt cagtgtgtcc caaggaaaac ccaaaagtga ggtgctcagg 840
tcagggggag tccagtgggc aggacccttg gcaggcaagc ccctcccttc actccaggga 900
cctaccttct gctagtaaag gactaggctt cattctaatt atggcccaca gactccccg 960
gagacctgga ggacagcagt gctggcaact ggggtgtccat gggcccgctc gccgctctg 1020
cctgtgctgc aagttttggc cgtgggtcca gccaaacaact ccctacgtcc tgtgtggggc 1080
cctggccaaag ttgatggggc attccttgag gagtatcatt ttcccagaca atccccatca 1140
cctttagggg ttctctgctt ggtctcttcc cagctgaaaa actagacctg tggcatggg 1200
gaagctggac aaagtctagg gggcccgctt ggtagaggtt cccgggaage ttgatctgtc 1260
agcctggccc ctgagccccc tgttaactca agactgtgag ctgctcttag gtggtcaact 1320
ctggagacta gcttgtatgg ctcttgacca gtatcaggat ttctgttctg agagcagct 1380
gggcagcaag gcaggcgagc ccagaggtgg cagcggcagg caatctgtgt actaggtctt 1440
tgtgatgcca aaaaataaag aggttggggg ggggtctttc tgttctctg attggatgga 1500
gtccgcagc aggtcagggc ctacattcca gtgctgact atagggagcc actcctgatt 1560
ccatggagca gcccgactt tgagaatggg ctctggtttg cggggggcag gcgtaccaga 1620
ctccaagacc ccccagtaac tcaccgtgcc aaataggaag aggtggcctt ggtgtagcca 1680

```

WO 02/18575

PCT/US01/26682

```

aatggatcct tttaacagtg tgcctttggg gagggaccga tgtccatggc ttcgttgagg 1740
gccatccata tgcagctgg gggccaagccc acagtggccc atgttggcty cagcaggaat 1800
gggtgccacc tcggcgaatt gaagggcctaa gattcccaga tagctaggcc cagagctgga 1860
agcagaacag aaggggaaga gctgctccca caggagaggg agagattcca gctcactgcy 1920
cagcctggga ggaggcgtgg atcctggcac gctgagcctc aggcaccaga ctccctgtgc 1980
tcgacagcaa agtcttgact ccttctgct gagcactgtg ctacctcac tgcctcaaag 2040
ccagactaac agctctccaa gcocttgggg tgaotcggct tccagagct gttggagaa 2100
tgagatgtc tgcctcctgc tgccttggga ggccagatcc ctcccagca gccaggtctc 2160
tccagaacct gattcgtgc ctttctgtt accagctact tcaatccaa agtttgaac 2220
tgcagatacc ttaactccag ccaacttggc tcttactgt gttgtgtgtt tttcctgggt 2280
cttcaagagc gtgtgcagg caagtcccgt cactgggaac tgcaccagat gctcagactt 2340
ggttgtctta tgtttaccaa taataaaag tagacttttt ctatttttat ttgctgetaa 2400
aaaaa 2405

```

```

<210> 8
<211> 2159
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 475113.7

```

```

<220>
<221> unsure
<222> 322-346
<223> a, t, c, g, or other

```

```

<400> 8
agagtcocgc cagcctcag agaattctgt gactgattcc aactccgatt cagaagatga 60
aagtggatg aattttctgg agaaaagggc ttaaatata aagcaaaaac aagcaatgct 120
tgcaaaactc atgtctgaat taaaagctct cctctgggaa ttccctggaa gaactccct 180
cccagctccc gactcacaat caagagacc ggaagcgtt acattccgg gtgttgcttc 240
caggagaaac cctgaacgga gagctcctcc tctaccagg tcaagctccc ggtactctcg 300
gtcccttgac gctctacca tnnnnnnnn nnnnnnnnn nnnnnntaca tgttggtgag 360
aagaggaag accgtggatg gctacatgaa tgaagatgac ctgccacgaa gccgtgctc 420
cagatcaacc gtgacccttc cgcataataat tgcocagtg gaagaaaita cagagagga 480
gttgagaaac gctgcagaca attctcagaa gaagatataat aaccgttcc tgggctctac 540
ttgtcatcaa tgcctcaga agactattga taccaaaaca aactgcagaa acccagactg 600
ctgggcgctt cagggccagt tctgtgccc ctgccttcca aaccgttatg gtgaagaggt 660
caggatgct ctgctggatc cgaactggca ttgcccgcct tctcagagaa tctgcaactg 720
cagtttctgc cggcagcag atggaaggtg tgcgactggg gtccttgtgt atttagccaa 780
alatacagc tttgggaatg tgcataccta ctgaaaagc ctgaaacag aatttgaat 840
gcaagcataa tatctggaaa atttctgccc tgcctctcac ttctcaaatc tttcttgtaa 900
aagtttccaa tttttctact gaacccctgag ttaaaaatct tgatgatcag cctgtttcat 960
aagaaactcc aatcaagtta atcttagcag acatgtgttt ctggagcacc acagaagtta 1020
tatgtctagt tacccttgg cctcctcagc tttctctctc gctcccaacc cccactctac 1080
agatccccc tctatttcca atgctcctct ccaacogctt agtttctgaa tttcttttaa 1140
attacagttt tatgaagaca tattttatct acttggtgtt gaaatagccc tcaaaaacc 1200
taagcacttg gaaaacacaa aatagtatta actaactaga tctattgaa ttcagagaag 1260
agcctctcaa ctgttttaca caaaaacgag tatgatttag cattcaact agttgaaatt 1320
ttaatagaa tcaaggoaca aaagtcttaa aaccatgtag aaaaattag taattattgc 1380
agattgatgt ctctcaatcc catgtattgc gcttatgtta caagttgtt tcaagttgaa 1440
gaottaatct ctctaatctt ctctcggccc aagggttaagt ggtgcccgtc agcttaaca 1500
atcacaatc aaggttgggt gggcaatgta atacttaatt aaaaataatg tggagagct 1560
atctggagat tatgagtag ctgattgaa tttccagat aaaaacttag tataattgta 1620
gtttgcaaa tttattccag ttcaactgta aggtattgca aataaattct tggacaattt 1680
tgaatggaaa ctgatatta aaaaactatc tgtggttctt tgcagttctc tgaataatta 1740
taaccaggcc acaaggttca agtttagatt ttaagcaact ttaatacaat gataagttcc 1800
tttttggaga tgtaactttt agcagttgtt taacctgaca tctctgccag tctagtttct 1860
gggcaggttt cctgtgtcag tattcccctc cctctttgca ttaatacaag tatttggtag 1920
aggtggatc taagtgtttg tatgtccaat ttacttgcac atgtaaaaca ttgctgtgcc 1980
atccaatgtt tgaatcataa ttggacctg aatcgataag tgtaaatata gcttttgatc 2040
tgaatgctt ttatacaaaa gtttatltta ataataaaat gtttgttcta acttctctgc 2100

```

WO 02/18575

PCT/US01/26682

ttttttaaaa ataactctac tgtacttaat tctaattttt tccatatt taaataaaa 2159

<210> 9
<211> 535
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 898622.1

<400> 9
cccaagtgc tgagattgca ggcgtgataa acaaatattc ttaatagggc tactttgaat 60
taactgcctc ttatgtttgg gagaagaaag ctgagacatt gcaatgaaaga tgatgagaga 120
taaatgttga tcttttggcc ccatttgtaa atgtattca gtatttgaac gtgctcctgt 180
ttatgttag ttttctcat catttattgt atagacaatt tttaaatctc tgtaatatga 240
taacttttcc tatcttttaa gttattgtaa cctaaagtta atccagatta tatggctcct 300
atatgtgtac aacattaaaa tgaaggctt tgtcttgcac tgtgaggtae agggggaagt 360
tggatcagg tttttaggact ctgtctctca ttagctgaaat aatgtgagga ttaactcttg 420
ccagctcaga ccatttctca atcagttgaa agggaaacaa gtattctagt ctcaaaattg 480
aataatgcac aagctttaag tgattaaaa aaaactgttc ttatgtcaaa aaaaa 535

<210> 10
<211> 2373
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 978267.1

<400> 10
ggttgactgt agagccgctc tctctcactg gcacagcgag gttttgctca gcccttgtct 60
gggagccgca ggtagctgtc tggcgacttc ttccgggtgt cccctccgc cctcctctgc 120
cctaccagct ttcttctctc cctgcccact ctcccgccgt ccccgagcc ccggccgagc 180
gccatggctc ctaggagggg cagtagtctg gtggccaaga ccaactcctt acggaggcgg 240
aagctcgctc cctttctgaa agactctgac cgtgaaatgg aaatacgaat caagcaaat 300
gagtcagaca gccacaactc cctcaaggag gtgataaacc tctacaacat cgagatcctg 360
ggctctccca aggtctctgc cgagatgaa tggcttgact acttcgccc ttggaggaaac 420
aaacaggccc tggagagggc ggcaacagct gacctggata tcaccgaaat aaacaaacta 480
acagcagaag ctattcagac acccctgaaa tctgcaaaaa cacgaaaggt aatacaggt 540
gatgaaatga tagtggaaga gggagaaggg agaagggaaa tttactgaag aatctctcaa 600
ctgcaagagc caaaaggtgt cctccatcca agaagagaac tcagttcata caaggcaag 660
gaaagggaaa aggtcaagc cgtgctaaca ctgttaccoc agctgtgggc cgtatggagg 720
tgtccatggt caaaocaaact ccaggcctga caccaggtt tagactcaag gtcttcaaga 780
ccctggcctg cgtactccag cagcaggaga gcggattac aacatctcag ggaatggcag 840
ccctctgtct gcacgcaagc agatcttctc cactgtgcca gtggccggcg gagagagcct 900
gcgattattg gccagtgact tgcagaggca cagtattgcc cagctggatc cagagggcct 960
gggaaacatc aagaagctct ccaaccgtct cgcccaaatc tgcagcagca tacggaccca 1020
caaatgagac accaaagttg acaggatgga cttttaatgg gcaactctgg gaccctgaag 1080
agactctctc cctcaggctc tattgttga gtgtgaagtt ccagagcaag gagccatgct 1140
cctcaaggg aattcaggaa ttcagagctg ctagtcccac accagttagg tagagctgtc 1200
tgttccacct ccatccacc ctgatccagc tcactgcttg ctggggccat gccatggaa 1260
cttccatca gtctccacc tgaatctctc ctgctctctg agctgtctgc tttgctctc 1320
tgcacatcaa catcctcttc accctgacct gcctgcagtt gagggggcga agaagaaccc 1380
tgtgttctca ggaagactgc ctccaccacc gatacccaaga gaactctgc atctggcatt 1440
tctgtctctc atgcttgaag ccggaggtt taggctcaga taagtgaact ctggccatg 1500
agagggtagg tccagaggtt ggggggaact gtacagatca gcagagcagg acagttggca 1560
gcagtgactc cagtagggaa catgtccctc tacctctctg cactatgac acctccctct 1620
accagctctc ctctctctca cctctctgt gggaggtggc cagtgggact tagggatctt 1680
tccctctctg tgcaccagtag ttctgaagtc tgcctgtgga cagtgctttt atgtttatcc 1740
ctgtttactg aagacaaact actggtttgg agacaacttc catgtctctc tctctactc 1800
ccctagttag tggaaatttg gataagggaa ctgtagggcc cagattctgg aggttttatg 1860
tcatgtgcca cagaataact gtctctaagc tatccatggt ccagtggtcc ctgccaagtc 1920

WO 02/18575

PCT/US01/26682

```
tgtagacttc agagagcact tototcttat ggggttcctg ggaacagggg cgggtgtgac 1980
ttgcttggtg gcctcattcc atgtgtgcct gtgcctgggg catggacttt gtaagcaga 2040
gtcagcagtg aggtectcat tctccagcca gccctctctc cctggagaat catgtgctat 2100
gttctaagaa tttgagaact agagtcctca tccccaggct tgaaggcaca tggctttctc 2160
atgtagggct ctctgtgta tttgttatta ttttgaaca agaccatttt agtaaacag 2220
tcctgttcaa gttgtattct ttaagttct ttattctctc ttccctgag atttttgat 2280
atatgtttct gagtaatggt atctttgagc tgattgttct aalcagagct ggtacctact 2340
ttcaataaat tctggttttg tgttttcttt tgt 2373
```

【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
7 March 2002 (07.03.2002)

PCT

(10) International Publication Number
WO 02/018575 A3

- (51) International Patent Classification: C12N 15/12, C07K 14/47, C12Q 1/68, A61K 38/17, G01N 33/68
- (21) International Application Number: PCT/US01/26682
- (22) International Filing Date: 27 August 2001 (27.08.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/229,253 30 August 2000 (30.08.2000) US
- (71) Applicant (for all designated States except US): INCVTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (72) Inventors; and
(73) Inventors/Applicants (for US only): WALKER, Michael, G. [CA/US]; 1050 Borregas Avenue, #80, Sunnyvale, CA 94089 (US); JUNG, Kenneth [US/US]; 725 Van Ness Avenue, #203, San Francisco, CA 94102 (US).
- (74) Agents: HAMLET-COX, Diana 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, GR, GU, HD, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MY, NA, NZ, PL, PT, RO, RU, SD, SI, SG, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SI, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IL, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CI, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NF, SN, TD, TG).
- Published:
with international search report
- (88) Date of publication of the international search report: 17 April 2003
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 02/018575 A3

(54) Title: GENES EXPRESSED IN THE CELL CYCLE

(57) Abstract: The invention provides cDNAs, their encoded proteins, and antibodies which may be used in methods for diagnosing and treating cell cycle disorders.

【国際調査報告】

INTERNATIONAL SEARCH REPORT		International Application No. PCT/US 01/26682
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/12 C07K14/47 C12Q1/68 A61K38/17 G01N33/68		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 7 C12N C07K A61K G01N		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the International search (name of data base and, where practical, search terms used) WPI Data, PAJ, CAB Data, SEQUENCE SEARCH, BIOSIS, EPO-Internal, GENSEQ		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	DATABASE EMBL-EBI [Online] European Bioinformatic Institute, Hinxton, UK; 9 March 2001 (2001-03-09) M.G. WALKER: "CDCA1 Cell division cycle associated gene 1, Human cDNA expression libraries, Homo sapiens cDNA clone 40371, mRNA sequence." Database accession no. BG354574 XP002203877 abstract --- -/--	1,6, 13-16,20
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "Z" document member of the same patent family		
Date of the actual completion of the international search 28 June 2002		Date of mailing of the International search report 21 10 2002
Name and mailing address of the ISA European Patent Office, P.O. 5818 Patentlaan 2 NL-2200 HV The Hague Tel: (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer HORNIG H.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/26682

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
L	DATABASE GENSEQ [Online] NCBI, Washington, US; 6 November 2001 (2001-11-06) Database accession no. AAK51682, AAK52666 XP002203878 * for the information: the accession nos. correspond to W00157190 * abstract	
P,X	-& WO 01 57190 A (HYSEQ, INC.) 9 August 2001 (2001-08-09) SEQ ID Nos. 227 and 2195; page 132 page 200; claims 1-28 ---	1,6, 13-16,20
L	DATABASE GENSEQ [Online] NCBI, Washington, US; 21 May 2002 (2002-05-21) Database accession no. ABK39802, ABK39803 XP002203879 * for the information: The accession nos. correspond to W00204514 * abstract	
E	-& WO 02 04514 A (CORIXA CORPORATION) 17 January 2002 (2002-01-17) SEQ ID No. 1931; claims 1-19 ---	1,6-13, 20
L	DATABASE GENSEQ [Online] NCBI, Washington, US; 12 March 2002 (2002-03-12) Database accession no. AAS99903 XP002203880 * for the information: the accession no. corresponds to W00185942 * abstract	
E	-& WO 01 85942 A (INCYTE GENOMICS INC.) 15 November 2001 (2001-11-15) SEQ ID Nos. 14 and 48 claims 1-44,58,92 ---	1-20
A	SAITO TADASHI ET AL: "Immunohistochemical analysis of cell cycle-associated proteins p16, pRb, p53, p27 and Ki-67 in oral cancer and precancerous carcinomas." reference to verrucous carcinomas." JOURNAL OF ORAL PATHOLOGY & MEDICINE, vol. 28, no. 5, May 1999 (1999-05), pages 226-232, XP001023157 ISSN: 0904-2512 the whole document --- -/--	

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/26682

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	HOLLYOAKE MARTINE ET AL: "The normal cell cycle activation program is exploited during the infection of quiescent B lymphocytes by Epstein-Barr virus." CANCER RESEARCH, vol. 55, no. 21, 1995, pages 4784-4787, XP002203876 ISSN: 0008-5472 the whole document	
A	SKOMEDAL HANNE ET AL: "Aberrant expression of the cell cycle associated proteins TP53, MDM2, p21, p27, cdk4, cyclin D1, RB, and EGFR in cervical carcinomas." GYNECOLOGIC ONCOLOGY, vol. 73, no. 2, May 1999 (1999-05), pages 223-228, XP001086409 ISSN: 0090-8258 the whole document	

INTERNATIONAL SEARCH REPORT	International application No. PCT/US 01/26682
Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)	
This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:	
1. <input type="checkbox"/> Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:	
2. <input type="checkbox"/> Claims Nos.: because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:	
3. <input type="checkbox"/> Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).	
Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)	
This International Searching Authority found multiple inventions in this International application, as follows:	
see additional sheet	
1. <input type="checkbox"/> As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.	
2. <input type="checkbox"/> As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.	
3. <input type="checkbox"/> As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:	
4. <input checked="" type="checkbox"/> No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:	
see invention 1. on extra sheet	
Remark on Protest	<input type="checkbox"/> The additional search fees were accompanied by the applicant's protest.
	<input type="checkbox"/> No protest accompanied the payment of additional search fees.

International Application No. PCT/US 01/26682

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: (1-20)-partially

A composition comprising a cDNA having the nucleic acid sequence of SEQ ID No. 1 or the complement thereof; a method for using said composition to detect gene expression in a sample; a method of using said composition to screen a plurality of molecules or compounds; a cDNA comprising a nucleic acid sequence selected from SEQ ID No. 1 and a complement thereof; a method for using said composition to detect gene expression in a sample; a method of using said composition to screen a plurality of molecules or compounds; an expression vector comprising a said cDNA selected from SEQ ID No. 1; a host cell comprising said vector; a method for using said SEQ ID No. 1 to produce a protein; a method for using said protein to screen a plurality of molecules or compounds to identify and purify at least one ligand which specifically binds the protein; a method of using said protein to prepare and purify antibodies;

2. Claims: (1-20)-partially

Idem as invention 1 but limited to SEQ ID No. 2;

3. Claims: (1-5)-partially

A composition comprising a cDNA having the nucleic acid sequence of SEQ ID No. 3 or the complement thereof; a method for using said composition to detect gene expression in a sample; a method of using said composition to screen a plurality of molecules or compounds;

4. Claims: (1-20)-partially

Idem as invention 1 but limited to SEQ ID No. 4;

5. Claims: (1-20)-partially

Idem as invention 1 but limited to SEQ ID No. 5;

6. Claims: (1-20)-partially

Idem as invention 1 but limited to SEQ ID No. 6;

7. Claims: (1-20)-partially

Idem as invention 1 but limited to SEQ ID No. 7;

International Application No. PCT/US 01/26682

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

8. Claims: (1-20)-partially

Idem as invention 1 but limited to SEQ ID No. 8;

9. Claims: (1-20)-partially

Idem as invention 1 but limited to SEQ ID No. 9;

10. Claims: (1-20)-partially

Idem as invention 1 but limited to SEQ ID No. 10;

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/26682

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0157190	A	09-08-2001	US 2002128187 A1 12-09-2002
			AU 2591801 A 31-07-2001
			AU 3128801 A 14-08-2001
			AU 3297101 A 07-08-2001
			AU 3300301 A 07-08-2001
			AU 3329301 A 14-08-2001
			AU 3484701 A 14-08-2001
			AU 3484801 A 14-08-2001
			AU 3486501 A 14-08-2001
			AU 3494401 A 14-08-2001
			AU 3665801 A 14-08-2001
			AU 3666001 A 14-08-2001
			AU 3666301 A 14-08-2001
			AU 3672101 A 14-08-2001
			AU 4314201 A 14-08-2001
			WO 0153326 A1 26-07-2001
			WO 0155334 A2 02-08-2001
			WO 0155335 A2 02-08-2001
			WO 0157255 A1 09-08-2001
			WO 0157260 A1 09-08-2001
			WO 0157175 A2 09-08-2001
			WO 0157261 A1 09-08-2001
			WO 0157262 A1 09-08-2001
			WO 0157187 A2 09-08-2001
			WO 0157265 A1 09-08-2001
			WO 0157188 A2 09-08-2001
			WO 0157266 A1 09-08-2001
			WO 0157267 A1 09-08-2001
			WO 0157190 A2 09-08-2001
			US 2002137044 A1 26-09-2002
			US 2002127199 A1 12-09-2002

WO 0204514	A	17-01-2002	AU 1877002 A 21-01-2002
			WO 0204514 A2 17-01-2002
-----	-----	-----	-----
WO 0185942	A	15-11-2001	AU 5945001 A 20-11-2001
			WO 0185942 A2 15-11-2001
-----	-----	-----	-----

フロントページの続き

(51) Int.Cl. ⁷	F I	テーマコード(参考)
A 6 1 P 13/12	A 6 1 P 17/00	4 C 0 8 4
A 6 1 P 17/00	A 6 1 P 19/02	4 H 0 4 5
A 6 1 P 19/02	A 6 1 P 19/10	
A 6 1 P 19/10	A 6 1 P 21/04	
A 6 1 P 21/04	A 6 1 P 25/00	
A 6 1 P 25/00	A 6 1 P 35/00	
A 6 1 P 35/00	A 6 1 P 37/00	
A 6 1 P 37/00	A 6 1 P 37/02	
A 6 1 P 37/02	C 0 7 K 14/47	
C 0 7 K 14/47	C 0 7 K 16/18	
C 0 7 K 16/18	C 1 2 N 1/15	
C 1 2 N 1/15	C 1 2 N 1/19	
C 1 2 N 1/19	C 1 2 N 1/21	
C 1 2 N 1/21	C 1 2 P 21/02	C
C 1 2 N 5/10	C 1 2 Q 1/68	A
C 1 2 P 21/02	G 0 1 N 33/15	Z
C 1 2 Q 1/68	G 0 1 N 33/50	Z
G 0 1 N 33/15	G 0 1 N 33/53	M
G 0 1 N 33/50	G 0 1 N 33/566	
G 0 1 N 33/53	C 1 2 N 5/00	A
G 0 1 N 33/566	A 6 1 K 37/02	

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

(72) 発明者 ジュング、ケネス

アメリカ合衆国カリフォルニア州 9 4 1 0 2 ・ サンフランシスコ ・ # 2 0 3 ・ バンネスアベニュー
7 2 5

F ターム(参考) 2G045 AA35 DA13 FB02 FB03
4B024 AA01 AA11 BA21 BA56 CA04 DA02 EA02 GA11 HA01 HA14
4B063 QA13 QQ43 QQ48 QR32 QR55 QS34
4B064 AG02 AG26 CA10 CA19 CC24 DA01 DA13
4B065 AA90X AA99Y AB01 AC14 BA02 CA24 CA44
4C084 AA01 AA07 BA01 NA14 ZA01 ZA45 ZA59 ZA66 ZA81 ZA89
ZA94 ZA96 ZA97 ZB05 ZB07 ZB08 ZB15 ZB26
4H045 AA10 AA11 AA30 BA10 CA41 DA01 EA28 EA51 FA74

专利名称(译)	在细胞周期中表达的基因		
公开(公告)号	JP2004521611A	公开(公告)日	2004-07-22
申请号	JP2002524078	申请日	2001-08-27
[标]申请(专利权)人(译)	洞察Genomics公司		
申请(专利权)人(译)	洞察基因组公司		
[标]发明人	ウォーカーマイケルジー ジュングケネス		
发明人	ウォーカー、マイケル・ジー ジュング、ケネス		
IPC分类号	G01N33/50 A61K38/00 A61P1/04 A61P9/10 A61P11/06 A61P13/12 A61P17/00 A61P19/02 A61P19/10 A61P21/04 A61P25/00 A61P35/00 A61P37/00 A61P37/02 C07K14/47 C07K16/18 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N15/09 C12N15/12 C12P21/02 C12Q1/68 G01N33/15 G01N33/53 G01N33/566		
CPC分类号	A61K38/00 A61P1/04 A61P11/06 A61P13/12 A61P17/00 A61P19/02 A61P19/10 A61P21/04 A61P25/00 C07K14/4738		
FI分类号	C12N15/00.ZNA.A A61P1/04 A61P9/10 A61P11/06 A61P13/12 A61P17/00 A61P19/02 A61P19/10 A61P21/04 A61P25/00 A61P35/00 A61P37/00 A61P37/02 C07K14/47 C07K16/18 C12N1/15 C12N1/19 C12N1/21 C12P21/02.C C12Q1/68.A G01N33/15.Z G01N33/50.Z G01N33/53.M G01N33/566 C12N5/00.A A61K37/02		
F-TERM分类号	2G045/AA35 2G045/DA13 2G045/FB02 2G045/FB03 4B024/AA01 4B024/AA11 4B024/BA21 4B024/BA56 4B024/CA04 4B024/DA02 4B024/EA02 4B024/GA11 4B024/HA01 4B024/HA14 4B063/QA13 4B063/QQ43 4B063/QQ48 4B063/QR32 4B063/QR55 4B063/QS34 4B064/AG02 4B064/AG26 4B064/CA10 4B064/CA19 4B064/CC24 4B064/DA01 4B064/DA13 4B065/AA90X 4B065/AA99Y 4B065/AB01 4B065/AC14 4B065/BA02 4B065/CA24 4B065/CA44 4C084/AA01 4C084/AA07 4C084/BA01 4C084/NA14 4C084/ZA01 4C084/ZA45 4C084/ZA59 4C084/ZA66 4C084/ZA81 4C084/ZA89 4C084/ZA94 4C084/ZA96 4C084/ZA97 4C084/ZB05 4C084/ZB07 4C084/ZB08 4C084/ZB15 4C084/ZB26 4H045/AA10 4H045/AA11 4H045/AA30 4H045/BA10 4H045/CA41 4H045/DA01 4H045/EA28 4H045/EA51 4H045/FA74		
优先权	60/229253 2000-08-30 US		
外部链接	Espacenet		

摘要(译)

本发明提供了编码哺乳动物GRIIP的哺乳动物cDNA。它还提供了cDNA，其片段，补体和变体以及所编码的蛋白质，其部分和其抗体的用途，用于诊断和治疗与炎症和免疫应答有关的疾病，特别是免疫系统的癌症。本发明还提供了用于产生蛋白质的表达载体和宿主细胞以及转基因模型系统。

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

