

(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2004-516815

(P2004-516815A)

(43) 公表日 平成16年6月10日(2004.6.10)

(51) Int. Cl. ⁷	F I	テーマコード (参考)
C 1 2 N 15/09	C 1 2 N 15/00 Z N A A	2 G O 4 5
A O 1 K 67/027	A O 1 K 67/027	4 B O 2 4
C O 7 K 14/47	C O 7 K 14/47	4 B O 2 9
C O 7 K 16/18	C O 7 K 16/18	4 B O 6 3
C 1 2 M 1/00	C 1 2 M 1/00 A	4 B O 6 4
	審査請求 未請求 予備審査請求 有 (全 101 頁) 最終頁に続く	

(21) 出願番号 特願2002-509376 (P2002-509376)
 (86) (22) 出願日 平成13年7月11日 (2001. 7. 11)
 (85) 翻訳文提出日 平成15年1月14日 (2003. 1. 14)
 (86) 国際出願番号 PCT/US2001/021982
 (87) 国際公開番号 W02002/004513
 (87) 国際公開日 平成14年1月17日 (2002. 1. 17)
 (31) 優先権主張番号 09/614, 474
 (32) 優先日 平成12年7月11日 (2000. 7. 11)
 (33) 優先権主張国 米国 (US)

(71) 出願人 301005050
 インサイト・ゲノミックス・インコーポレ
 イテッド
 アメリカ合衆国カリフォルニア州9430
 4・パロアルト・ポータードライブ 31
 60
 (74) 代理人 100089266
 弁理士 大島 陽一
 (72) 発明者 ローリング、ジーン・エフ
 アメリカ合衆国カリフォルニア州9440
 4・フォスターシティ・サンタクルスレ
 ン 701

最終頁に続く

(54) 【発明の名称】 ダウン症候群重要領域1-様1タンパク質

(57) 【要約】

本発明は、核酸分子及びその断片を提供する。本発明はまた、この核酸分子を用いた遺伝子発現に関連する症状、疾患、及び障害の特徴付け・診断・評価・治療方法、並びにモデル系の作製方法を提供する。更に本発明は、この核酸分子によってコードされるタンパク質を産生するための発現ベクター及び宿主細胞を提供する。

```

190      199      238      217
5'  ATG TCA ACT TTG GAA AAA AAC AAC GTA GTG TTG
      M S T L R K N N V V X
226      235      244      253      262      271
GGA ATG AGG GGA GAA TCA TAC TTC ATC GGA ATG ACG AGC CCA GGC CAG CAG GGA
      G M R G E L Y F I C N R S P G Q Q G
280      283      298      307      316      325
CAC GTC CCT GAA GAT GAA GGA CTT TTC TTA CTG TGC TGC ATA GAC AGG GAC TGG
      H V P E D G G L F L L C C I D P D W
334      343      352      361      370      379
GCT GTC ACT CGV NUT TTT GCA GAA GAA GGC TTT CAA GGA ATC ACT GAC TTC AAT
      A V T R C F A E E A F Q A I T D F N
388      397      406      415      424      433
GAC CTC CCC AAC TCG TTG TTF GCG TGC AAT GTT CAC CAG TCA CTG TTT GAA GGA
      D L P N S L P A C M V H Q S V P E G
442      451      460      469      478      487
GAA GAG AGC AAG GAA AAA TTT GAG GGA CTG TTT CCG ACT TAT GAT GAC TGT CTG
      E E S K E K F E G L F R T Y D D C V
    
```

【特許請求の範囲】

【請求項 1】

(a) SEQ ID NO: 2 のアミノ酸配列、
 (b) SEQ ID NO: 2 の抗原エピトープ、及び
 (c) SEQ ID NO: 2 のアミノ酸配列と少なくとも 95% の同一性を有する変異体から選択されるタンパク質またはその一部をコードする単離された核酸分子またはその断片。

【請求項 2】

(a) SEQ ID NO: 1 の核酸配列、
 (b) SEQ ID NO: 1 の核酸配列と少なくとも 85% の同一性を有する cDNA、
 及び
 (c) SEQ ID NO: 1 の約 1 ~ 400 ヌクレオチドの SEQ ID NO: 1 の断片から選択される単離された核酸分子またはその相補体。

【請求項 3】

SEQ ID NO: 3 - 9 から選択される請求項 2 の cDNA。

【請求項 4】

請求項 1 の核酸分子またはその相補体とレポーター分子を含む組成物。

【請求項 5】

請求項 1 の核酸分子またはその相補体を含む基板。

【請求項 6】

請求項 1 の核酸分子またはその相補体を含むアレイ。

【請求項 7】

請求項 1 の核酸分子を含むベクター。

【請求項 8】

請求項 7 のベクターを含む宿主細胞。

【請求項 9】

タンパク質を生産するための方法であって、

(a) 前記タンパク質が発現する条件下で、請求項 8 の宿主細胞を培養するステップと、
 (b) 前記宿主細胞から前記タンパク質を回収するステップとを含むことを特徴とする方法。

【請求項 10】

請求項 7 のベクターを含む遺伝子組換え細胞系または生物。

【請求項 11】

核酸を含むサンプルの核酸分子の発現変動 (differential expression) を検出する方法であって、

(a) 請求項 1 の核酸分子と前記サンプルの核酸とをハイブリダイズさせて、ハイブリダイゼーション複合体を形成するステップと、
 (b) 前記サンプルにおける前記核酸分子の存在を表す前記ハイブリダイゼーション複合体を定量するステップとを含むことを特徴とする検出方法。

【請求項 12】

更に、前記ハイブリダイゼーションの前に前記サンプルの核酸を増幅するステップを含むことを特徴とする請求項 11 に記載の方法。

【請求項 13】

発現変動の検出によりアルツハイマー病、ダウン症候群、または痴呆の他の形態の診断ができることを特徴とする請求項 11 に記載の方法。

【請求項 14】

核酸分子を用いて複数の分子または化合物をスクリーニングする方法であって、

(a) 結合が許容される条件下で、請求項 1 の核酸分子を複数の分子または化合物と結合させるステップと、
 (b) その結合を検出して、前記核酸分子に結合する分子または化合物を特定するステッ

10

20

30

40

50

ブとを含むことを特徴とするスクリーニング法。

【請求項 15】

前記分子または化合物が、DNA分子、RNA分子、ペプチド核酸、人工染色体作製物、ペプチド、転写因子、リプレッサー、及び調節分子から選択されることを特徴とする請求項 14 に記載の方法。

【請求項 16】

(a) SEQ ID NO: 2 のアミノ酸配列、
(b) SEQ ID NO: 2 の抗原エピトープ、及び
(c) SEQ ID NO: 2 のアミノ酸配列と少なくとも 95% の同一性を有する変異体から選択されることを特徴とする精製されたタンパク質またはその一部。

10

【請求項 17】

請求項 16 のタンパク質及びレポーター分子を含む組成物。

【請求項 18】

タンパク質を用いて、複数の分子または化合物をスクリーニングして少なくとも 1 つのリガンドを特定する方法であって、

(a) 特異的な結合が許容される条件下で、請求項 16 のタンパク質を複数の分子または化合物と結合させるステップと、
(b) 特異的な結合を検出して、前記タンパク質に特異的に結合するリガンドを特定するステップとを含むことを特徴とする方法。

【請求項 19】

タンパク質を用いて抗体を生産及び精製する方法であって、

(a) 動物を請求項 16 のタンパク質で免疫して、抗体反応を生じさせるステップと、
(b) 動物抗体を単離するステップと、
(c) 前記タンパク質を基板に付着させるステップと、
(d) 前記タンパク質に対する特異的な結合が許容される条件下で、単離した前記抗体に前記基板を接触させるステップと、
(e) 前記タンパク質から抗体を解離させて、前記タンパク質と特異的に結合する精製された抗体を得るステップとを含むことを特徴とする方法。

20

【請求項 20】

請求項 16 のタンパク質と結合する抗体。

30

【発明の詳細な説明】

【0001】

(発明の技術分野)

本発明は、新規の哺乳動物タンパク質をコードする核酸分子に関し、また、これらの分子を用いたアルツハイマー病、ダウン症候群、または痴呆の他の形態などの症状の特徴付け・診断・治療に関する。

【0002】

(発明の背景)

アルツハイマー病は、アミロイド ペプチドを含む老人斑及び神経原線維濃縮体の形成によって特徴付けられる進行性の神経変性疾患である。このような老人斑は、海馬、前頭皮質、帯状皮質、小脳、基底核、及び青斑核を含む脳の連合皮質及び辺縁に見られる。初期のアルツハイマー病では、帯状皮質の物理的変化が視認できる (Minoshima et al. (1997) *Annals of Neurology* 42: 85-94)。後期のアルツハイマー病患者では、蓄積された老人斑により辺縁領域の神経構造が破壊され、最終的に記憶過程に障害が起こる。

40

【0003】

全世界には、約 2000 万人ものアルツハイマー病による痴呆患者がいる。この疾患は、早期発症型 (30 歳の若者も発病する)、家族性、または散発性の場合がある。家族性アルツハイマー病はかつて、完全に常染色体優性な特性であると考えられていたが、この考えは多くの遺伝的決定因子が排除されていくにつれて変化してきた。例えば、老人斑に見

50

られるアポリポタンパク質 E (Apo E) の或る正常なアレル変異体は、アルツハイマー病を発症するリスクを増大し得るか、或いはアルツハイマー病の発症を防止し得るかの何れかである (Strittmatter et al . (1993) Proc Natl Acad Sci 90 : 1977 - 1981) 。

【 0004 】

Apo E、アミロイド前駆体タンパク質 (APP)、プレセニン - 1、及びプレセニン - 2 の 4 種の遺伝子における突然変異により、アルツハイマー病に罹患し易くなることが知られている (Selkoe (1999) Nature 399 : A23 - A31) 。 Apo E 遺伝子の e4 アレルが、晩年発症型アルツハイマー病のリスクを増大させる。

アミロイドタンパク質 (A) が老人斑の主成分であり、この A は通常、セクターゼ及びセクターゼにより APP が切断されて形成される。アルツハイマー病患者では、多量の A がこのような神経病理学的な斑で生成され細胞外に蓄積される。

【 0005 】

アルツハイマー病と多くの他の遺伝子やタンパク質との関連が報告されている。胎児アルツハイマー抗原及び synuclein が、脳の斑及び神経原線維濃縮体で見られる。或る種の遺伝子多形の遺伝もまた、アルツハイマー病のリスク増大に関連する。例えば、プロテアーゼインヒビターとして機能し得る 2 - マクログロブリンをコードする或る種の遺伝子多形が、晩年発症型アルツハイマー病のリスク増大に係る。

【 0006 】

マイクロアレイ技術を用いた実験により、アルツハイマー病患者の組織において、特定の遺伝子の発現が変化することが分かった。このような遺伝子の 1 つである DSCR1L1 (g1435040) と呼ばれるダウン症候群重要領域 1 - 様 1 (Mazowiecki et al . (1996) J Biol Chem 271 : 14567 - 14571) が、アルツハイマー病の組織において 2 分の 1 以下にダウンレギュレートされることが分かった。DSCR1L1 は、甲状腺ホルモン応答性タンパク質をコードし、DSCR1 (g7657042) 及び DSCR1L2 (g6017918) を含む遺伝子ファミリーのメンバーである。このファミリーの第 1 のメンバー DSCR1 がクローニングされ、第 21 染色体のダウン症候群領域に近接していることから命名されたが、そのコードされたタンパク質の機能及びダウン症候群におけるその役割は不明である。このファミリーのタンパク質であることを決定するモチーフには、多くの RNA 結合タンパク質及び或る種の一本鎖 DNA 結合タンパク質に見られるドメインに類似した N 末端 RNA 結合ドメインと、リン酸化の標的と思われる ISPPXSP ボックスを含む中心に存在する短いユニークなセリン - プロリンモチーフとが含まれる (Fuentes et al . (1995) Hum Mol Genet 4 : 1935 - 1944) 。

【 0007 】

アミノ酸配列の一般的な特性から判断すると、DSCR1 遺伝子が転写調節及びシグナル伝達に参与するタンパク質をコードする可能性が高い (Strippoli et al . (2000) Genomics 64 : 252 - 263) 。 3 種類の遺伝子、即ち DSCR1、DSCR1L1、及び DSCR1L2 が発現変動するのが明らかである。DSCR1 は胎児の脳及び心臓において多量に発現し、DSCR1L1 は心臓、脳、肝臓、及び筋骨格で見られ、DSCR1L2 は血液を含む殆どの組織で発現する。

【 0008 】

アルツハイマー病とダウン症候群との間には、中年のダウン症候群患者に見られるアミロイドの重度の蓄積などの幾つかの関連性が存在する。DSCR1L1 の甲状腺ホルモン応答特性が、アルツハイマー病及び / またはダウン症候群におけるこの遺伝子の役割に関連し得る。すなわち、アルツハイマー病及びダウン症候群患者の双方において、アルツハイマー病患者の脳脊髄液における抗チログロブリン抗体の上昇及びプレアルブミンのレベルの低下、並びにアルツハイマー病の臨床症状のあるダウン症候群患者における明白な無症状の甲状腺機能低下を含む甲状腺ホルモンの代謝障害が見られる (Sutherland et al (1992) Neurobiol Aging 13 : 301 - 312) 。

これとは逆に、甲状腺機能が低下した患者には、アルツハイマー病に見られるのと同様の或る種の神経性症状が見られ、甲状腺機能不全の病歴が、アルツハイマー病を発症するリスク因子であると考えられる。考えられる得る1つの関連性は、甲状腺ホルモンがAPP遺伝子の転写活性を負に調節するという発見である (Beland et al. (1998) J Biol Chem 273: 30366-30371)。APPの過剰な発現により、アミロイドタンパク質の産生増大を伴うメカニズムによって神経変性が引き起こされるとの仮説がある。アルツハイマー病患者における甲状腺応答DSCR1L1遺伝子のダウンレギュレーションと他のDSCR1遺伝子のダウンレギュレーションの可能性とから、アルツハイマー病と甲状腺機能との間に関連があることが分かる。

【0009】

10

DSCR1タンパク質ファミリーのメンバーをコードする更なる核酸分子の発見を、アルツハイマー病、ダウン症候群、及び他の痴呆の形態の診断・予後・治療に利用することができる。

【0010】

(発明の概要)

本発明は、アルツハイマー病、ダウン症候群、または痴呆の他の形態などの症状の特徴付け・診断・予防・治療に有用な組成物を提供することで当分野の要望を満たすDSCR1L1タンパク質をコードする精製された核酸分子の発見に基づいている。

【0011】

本発明は、SEQ ID NO: 2のアミノ酸配列、SEQ ID NO: 2のアミノ酸配列と少なくとも95%の同一性を有する変異体、及びSEQ ID NO: 2の抗原性エピトープを含む群から選択されるタンパク質またはその一部をコードする単離された核酸分子またはその断片を提供する。本発明はまた、SEQ ID NO: 1の核酸配列、SEQ ID NO: 1の核酸配列と少なくとも85%の同一性を有する変異体cDNA、及びSEQ ID NO: 1の断片を含む群から選択される単離された哺乳動物核酸分子またはそのその相補体を提供する。更に本発明は、DSCR1L1をコードする核酸分子またはその相補体を含む組成物、基板、及びプローブを提供する。更に本発明は、このような核酸分子を含むベクター、このようなベクターを含む宿主細胞、及びこのような核酸分子を用いてDSCR1L1を作製する方法を提供する。更に本発明は、DSCR1L1をコードする核酸分子を含むベクターを有する遺伝子組換え細胞系または遺伝子組換え生物を提供する。本発明はまた、SEQ ID NO: 3-9から成る群から選択される哺乳動物断片またはその相補体を提供する。或る実施態様では、本発明は核酸分子を含む基板を提供する。第2の実施態様では、この基板を、検出法、スクリーニング法、及び精製法に用いる。更なる実施態様では、核酸分子は一本鎖相補RNAまたはDNA分子である。

20

30

【0012】

本発明はまた、核酸分子を用いて、サンプルの核酸の発現変動を検出する方法であって、前記核酸分子と前記核酸とをハイブリダイズさせてハイブリダイゼーション複合体を形成し、そのハイブリダイゼーション複合体を標準と比較し、その比較によりサンプルにおける前記核酸の発現変動が分かる方法を提供する。或る実施態様では、この検出方法は、ハイブリダイゼーションの前に前記核酸を増幅するステップを含む。別の実施態様では、前記核酸分子の発現変動を標準と比較して、アルツハイマー病、ダウン症候群、及び痴呆の他の形態の診断をする。更なる実施態様では、前記核酸分子、その断片やその相補体はアレイ上のエレメントを含み得る。

40

【0013】

本発明はまた、核酸配分子、その断片またはその相補体を用いて、ライブラリ即ち複数の分子または化合物をスクリーニングしてその核酸分子と特異的に結合する少なくとも1つのリガンドを同定する方法であって、特異的な結合が許容される条件下で、その核酸分子を前記分子または化合物と結合させるステップと、前記核酸分子に対する特異的な結合を検出して、前記核酸分子と特異的に結合するリガンドを特定するステップとを含む方法を提供する。一実施態様では、このような分子または化合物として、DNA分子、RNA分

50

子、ペプチド核酸、人工染色体作製物、ペプチド、転写因子、リプレッサー、及び調節分子が選択される。

【0014】

本発明はまた、本核酸分子の少なくともある断片を含む発現ベクターを提供する。別の実施態様では、この発現ベクターは宿主細胞内に含まれている。本発明は更に、タンパク質が発現する条件下で宿主細胞を培養するステップと、そのタンパク質をその宿主細胞から回収するステップとを含むタンパク質生産方法を提供する。

【0015】

本発明はまた、SEQ ID NO: 2のアミノ酸配列と、SEQ ID NO: 2のアミノ酸配列と95%の同一性を有する変異体と、SEQ ID NO: 2の抗原エピトープと、SEQ ID NO: 2のオリゴペプチドと、SEQ ID NO: 2の生物学的に活性な部分とからなる群から選択される精製されたタンパク質またはその一部を提供する。本発明はまた、前記精製されたタンパク質及び医薬用担体を含む医薬組成物を提供する。本発明は更に、タンパク質を用いてライブラリ即ち複数の分子または化合物をスクリーニングして少なくとも1つのリガンドを特定する方法であって、特異的な結合が許容される条件下で、そのタンパク質を前記分子または化合物と結合させるステップと、特異的な結合を検出して、前記タンパク質と特異的に結合するリガンドを特定するステップとを含む方法を提供する。ある実施態様においては、分子または化合物はDNA分子、RNA分子、ペプチド核酸、ペプチド、タンパク質、擬態、アゴニスト、アンタゴニスト、抗体、免疫グロブリン、インヒビター、及び薬剤から選択される。別の実施態様では、アルツハイマー病、ダウン症候群、または痴呆の他の形態の患者を治療するために前記リガンドを使用する。

10

20

【0016】

本発明は、哺乳動物タンパク質を用いて、被検サンプルをそのタンパク質に特異的に結合する抗体に対してスクリーニングする方法であって、前記被検サンプルから抗体を単離するステップと、特異的な結合が許容される条件下で、この単離した抗体を前記タンパク質とを結合させるステップと、結合した前記タンパク質から抗体を解離するステップと、抗体の存在或いは存在量からアルツハイマー病、ダウン症候群、または痴呆の他の形態の患者を診断できる抗体の量を既知の標準値と比較するステップとを含むスクリーニング法を提供する。本発明はまた、哺乳動物タンパク質を用いて抗体を作製及び精製する方法であって、抗体反応が起こる条件下でこのタンパク質によって動物を免疫するステップと、動物抗体を単離するステップと、基板にこのタンパク質を付着させるステップと、このタンパク質への特異的な結合が許容される条件下で、この基板を単離した抗体と接触させること、このタンパク質から抗体を分離して精製した抗体を得ることを含む方法を提供する。

30

【0017】

本発明は、DSCR1L1 に特異的に結合する精製された抗体を提供する。本発明はまた、抗体を用いてアルツハイマー病、ダウン症候群、または痴呆の他の形態を診断する方法であって、結合が許容される条件下で、サンプルと前記抗体とを結合させるステップと、結合した抗体の量を既知の標準値と比較して前記疾患の有無を診断するステップとを含む診断方法を提供する。本発明は更に、抗体を用いてアルツハイマー病、ダウン症候群、または痴呆の他の形態を治療する方法であって、精製された抗体を含む医薬組成物をそのような治療が必要な患者に投与することを含む治療方法を提供する。

40

【0018】

本発明はまた、哺乳動物のゲノムDNAの中にマーカー遺伝子を挿入して、内在性ポリヌクレオチドの発現を阻害する方法を提供する。本発明はまた、核酸分子を用いて哺乳動物モデル系を作製する方法であって、SEQ ID NO: 3-9から選択される核酸分子を含むベクターを作製するステップと、そのベクターで胚性幹細胞を形質転換するステップと、形質転換された胚性幹細胞を選択するステップと、この形質転換された胚性幹細胞を哺乳動物胚盤胞の中に微量注入して、キメラ胚盤胞を形成するステップと、偽妊娠メスにキメラ胚盤胞を導入し、このメスが、生殖細胞系の中に核酸分子を含むキメラ子孫を出産

50

し、そのキメラ哺乳動物を交配して同型接合哺乳動物モデル系を作製するステップとを含む哺乳動物モデル系作製方法を提供する。

【0019】

(本発明の実施方法について)

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

10

【0020】

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

【0021】

(定義)

「ダウン症候群重要領域1-様1(DSCR1L1)タンパク質」は、天然、合成、半合成或いは組換え体など任意の哺乳動物種(ウシ、イヌ、マウス、ヒツジ、ブタ、げっ歯類、サルそして好ましくはヒトを含む)から得られる実質的に精製されたタンパク質を指す。

20

【0022】

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

【0023】

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

30

【0024】

「核酸分子」とは、単離したポリヌクレオチド、cDNA或いはその任意の断片または相補配列を指す。それは組換えまたは合成された二本鎖または一本鎖であり、コード配列及び/または非移動配列(nonmoving sequence)、ゲノムDNA分子由来のイントロンを有するまたは有しないエキソンを有する。

【0025】

「タンパク質をコードする核酸分子」という語は、当分野で周知の分析により同定された保存された領域、モチーフ或いはドメインをコードする配列と密接にアラインメントする核酸配列を指す。これらの分析には、保存された領域内における同一性を特定するBLAST(Basic Local Alignment Search Tool)が含まれる(Altschul(1993) J Mol Evol 36: 290-300、Altschulら(1990) J Mol Biol 215: 403-410)。Brenner et al.(1998; Proc Natl Acad Sci 95: 6073-6078)が配列同一性による構造的相同性を特定するべくBLASTの信頼性を分析し、30%の同一性が少なくとも150残基の配列アラインメントに対する信頼できる閾値であり、40%の同一性が少なくとも70残基のアラインメントに対する適切な閾値であることを示した(Brenner et al., page 6076, column 2)。

40

50

【0026】

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

【0027】

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

10

【0028】

「障害」とは、核酸分子及びDSCR1タンパク質が異なって発現されている場合の症状、疾患、または症候群を指す。

【0029】

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

20

【0030】

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

【0031】

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

30

【0032】

「オリゴヌクレオチド」とは、長さが約18から約60ヌクレオチドの一本鎖分子を指す。そしてハイブリダイゼーションや増幅技術において、または複製、転写、翻訳の調節において使用され得る。同義語として、アンプライマー、プライマー、及びオリゴマーが挙げられる。

【0033】

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

【0034】

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

40

【0035】

「プローブ」とは、サンプル中の少なくとも1つの核酸にハイブリダイズする核酸分子を指す。標的が一本鎖である場合、プローブは相補的一本鎖である。プローブは、サザン法、ノーザン法、in situ法、ドットプロット法、及びアレイなどを含むハイブリダイゼーション反応またはスクリーニングアッセイで使用するためにレポーター分子で標識化することができる。

50

【0036】

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

【0037】

「精製された」とは、自然環境から分離され、自然環境で会合していた他の化合物が約60%から約90%まで分離したあらゆる分子や化合物を目指す。 10

【0038】

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

【0039】

「特異的な結合」とは、構造、特に分子側鎖に依存する2つの分子間での特殊な正確な相互作用を目指す。例えば、調節タンパク質のDNA分子の主溝への挿入、2つの一本鎖核酸間の骨格に沿った水素結合、またはタンパク質のエピトープとアゴニスト、アンタゴニストまたは抗体との間の結合がある。 20

【0040】

配列に適用される「類似性」とは、Smith-Waterman アルゴリズム(Smith及びWaterman(1981) J Mol Biol 147:195-197)或いはBLAST2(Altschulら(1997) Nucleic Acids Res 25:3389-3402)などの標準化されたアルゴリズムを用いてアラインメントされた少なくとも2つの配列間で一致する分子或いは残基の定量化(通常は%)を目指す。BLAST2は、アラインメントを最適化するために配列の1つでギャップを挿入するまた2つの配列をより有意に比較できる標準化された再現性のある方法で使用され得る。 30

【0041】

「基板」とは、核酸分子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目指すものであって、膜、フィルタ、チップ、スライド、ウエハ、ファイバー、磁性または非磁性ビーズ、ゲル、毛管、またはその他のチューブ、プレート、ポリマー、微小粒子が含まれ、穴、溝、ピン、チャンネル、孔等、様々な表面形態を有する。

【0042】

「変異体」とは、核酸分子またはその核酸分子がコードするタンパク質の認識できる変異した分子を目指す。スプライス変異体は、スコアが少なくとも100、最も好ましいのは少なくとも400であるBLASTスコアにより決定し得る。対立遺伝子変異体は、核酸分子に対して高い同一性のパーセントを有し、100塩基に付き約3塩基が異なり得る。「一塩基多型」(SNP)とは、欠失、挿入、または置換による単一塩基による変異を目指す。この変異は、保存的(プリンからプリン)或いは非保存的(プリンからピリミジン)であり得、コードされたアミノ酸に変異が起こる可能性がある。 40

【0043】

(発明)

本発明は、DSCR1L1 をコードする新規の哺乳動物核酸分子の発見に基づき、アルツハイマー病、ダウン症候群、及び他の形態の痴呆などの病態の特徴付け・診断・治療・予防においてその核酸分子やその断片、またはそのタンパク質やその一部を用いることに関連する。

【0044】

一実施形態では、本発明は図1A - 図1Cに示されているSEQ ID NO: 2のアミノ酸配列を含むポリペプチドを包含する。DSCR1L1は、既知のRNA認識モチーフに類似した残基129~199のプラスに荷電した芳香族アミノ酸によって特徴付けられる一本鎖核酸結合ドメインを有する。加えて、DSCR1L1は、DSCR1タンパク質の特徴であるISPPXSPボックス(残基169~176)を含む。図2A及び図2Bに示されているように、DSCR1L1は、ヒトDSCR1L1(g1435040; SEQ ID NO: 10)及びヒトDSCR1L2(g6017919; SEQ ID NO: 11)と化学的及び構造的相同性を有する。具体的には、DSCR1L1はDSCR1L1との同一性が90.3%であり、DSCR1L2とは57.3%の同一性を有する。これら3種類全てのタンパク質が、一本鎖核酸分子結合ドメイン及びISPPXSPボックスを共に有する。SEQ ID NO: 1は、5'非翻訳領域がDSCR1L1(g1435040; SEQ ID NO: 10)をコードする遺伝子及びDSCR1L2(g6017919; SEQ ID NO: 11)をコードする遺伝子とは異なり、また、SEQ ID NO: 1の読み枠によってコードされた残基M1からS70までの初めの70個のアミノ酸により、DSCR1L1ファミリータンパク質をコードするこれら及び他の遺伝子と区別される。SEQ ID NO: 1のヌクレオチド1から400までのプローブは、DSCR1L1をコードする自然発生の核酸分子、アレル変異体、または関連分子を特定するのに有用であろう。DSCR1L1のアミノ酸残基M1からS70までの抗原エピトープは、DSCR1L1と類似タンパク質とを区別し得るDSCR1L1の抗体の生産に有用であろう。

【0045】

電子ノーザン分析により、様々なライブラリにおいてこの配列の発現が示され、精神分裂病、ハンチントン病、癲癇、及び筋萎縮性側索硬化症に関連する組織を含む神経系の組織に最も多量に存在することが分かった。アルツハイマー病と診断された患者のサンプルとのマイクロアレイにおけるハイブリダイズが2分の1にダウンレギュレートされたのに一致して、アルツハイマー病の患者からのライブラリ8の内7のライブラリでDSCR1L1が発現していない。

【0046】

表1は、ラット及びサルからの核酸断片、その配列の範囲、及びSEQ ID NO: 1との同一性を示す。列1及び列2にはそれぞれ、各核酸配列に対する配列番号(SEQ ID NO)及びインサイト配列識別子が列記されている。列3には各断片のヌクレオチド長、列4には各断片がSEQ ID NO: 1と同一性を有するSEQ ID NO: 1におけるヌクレオチド残基の範囲、列5には由来する生物名、列6には列4のヌクレオチド範囲に対する各断片とSEQ ID NO: 1との同一性(%)が列記されている。

【0047】

これらの核酸分子は、ヒトの疾患モデルとなる遺伝子組換え生物または遺伝子組換え細胞系を作製するのに特に有用である。また、このような遺伝子組換え生物において治療の可能性を検査することができる。初期のアルツハイマー病の機能欠損を示す領域である帯状皮質におけるSEQ ID NO: 9の発現に注目されたい。このサル核酸分子は、アルツハイマー病のモデル系に有用であろう。また、サル核酸分子をプローブとして用いて、同様に発現が帯状皮質に制限される相同なヒトDSCR1遺伝子ファミリー配列を検索することができる。

【0048】

(本発明の特徴及び使用)

cDNAライブラリ

ここに開示する特定の実施例では、当分野で周知の方法を用いて哺乳動物の細胞及び組織からmRNAを単離し、これを用いてcDNAライブラリを作製する。上記したインサイトクローンは、哺乳動物cDNAライブラリから単離された。本発明の代表的な少なくとも1つのライブラリの作製方法を以降に記載する実施例に示す。コンセンサス哺乳動物

配列は、Phrap (P. Green, 前出) 及び GELVIEW 断片構築システム (Genetics Computer Group, Madison WI)、AUTO ASSEMBLER アプリケーション (Applied Biosystems, Foster City CA) などのコンピュータプログラムを用いて、インサイト社クローンを含む断片、伸長、及び/またはショットガン配列から化学的かつ/または電子的に構築した。

【0049】

シークエンシング

核酸をシークエンシングする方法は当分野で周知であり、そのような方法を用いて本発明の任意の実施例を実施することができる。これらの方法は、DNAポリメラーゼ I であるクレノウフラグメント、SEQUENASE、Taq DNAポリメラーゼおよび熱耐性 T7 DNAポリメラーゼ (Amersham Pharmacia Biotech (APB), Picataway NJ)、或いは ELONGASE 増幅システム (Life Technologies, Rockville MD) に用いられるような校正エクソヌクレアーゼとポリメラーゼとの組み合わせを用いることができる。配列の準備は、HYDRA マイクロディスペンサー (Robbins Scientific, Sunnyvale CA)、MICROLAB 2200 (Hamilton, Reno NV)、及び DNA ENGINE サーマルサイクラー (PTC 200; MJ Research, Watertown MA) などの装置を用いて自動的に行うのが望ましい。シークエンシングに用いる装置には、ABI 3700、377 または 373 DNA シークエンシングシステム (Applied Biosystems)、及び MEGABACE 1000 DNA シークエンシングシステム (APB) 等がある。当分野で周知の様々なアルゴリズムを用いて、シークエンシングした配列を解析することができる。これらのアルゴリズムは、Ausubel (1997; Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7) 及び Meyers (1995; Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856 - 853) に記載されている。

【0050】

ショットガンシークエンシングを用いて、様々な由来するクローニングしたインサートから更に配列を作り出す。ショットガンシークエンシング方法は当分野で周知であり、熱耐性 DNA ポリメラーゼや非熱耐性 DNA ポリメラーゼ、及び目的の核酸分子に隣接する代表的な領域から選択されたプライマーを用いる。当分野で周知の CONSED (Gordon (1998) Genome Res. 8: 195 - 202) などの様々なアルゴリズムやプログラムを用いて、組み立てが未終了の配列 (組み立てが不完全な配列) を調べる。ベクターやキメラ配列、または欠失配列を含む汚染配列を除去して、組み立てが未終了の配列を完全な配列に組み立てる。

【0051】

核酸配列の伸長

本発明の配列は、当分野で周知の様々な PCR 法を用いた方法で伸長することができる。例えば、XL-PCR キット (Applied Biosystems) 及び入れ子プライマー (nested primer)、市販の cDNA またはゲノム DNA ライブラリ (Life Technologies; Clontech, Palo Alto CA, respectively) 用いてヌクレオチド配列を伸長することが可能である。全ての PCR 系の方法に用いることができるように、プライマーは、OLIGO 4.0 6 プライマー分析ソフトウェア (Molecular Biology Insights, Cascade CO) 等の市販のソフトウェアを用いて、ヌクレオチドの長さが約 22 ~ 30 個、GC 含量が約 50% 以上、約 55 ~ 68 の温度で標的配列とアニールするように設計することが可能である。調節エレメントを復活させるために配列を伸長する場合は、cDNA ライブラリよりゲノムライブラリを用いる方が良い。

10

20

30

40

50

【0052】

(本哺乳動物核酸分子の使用)

ハイブリダイゼーション

本哺乳動物核酸分子及びその断片は、様々な目的のための様々なハイブリダイゼーション技術に用いることができる。プローブは、5'調節領域やDSCR1タンパク質ファミリーに見られるN末端RNA結合ドメインなどのユニークな領域から作製可能であり、これらのプローブを、この哺乳動物タンパク質やアレル変位配列、または関連分子をコードする天然の分子を同定するためのプロトコルに用いることができる。通常は一本鎖であるDNA若しくはRNAからなるこのプローブは、任意のこの核酸配列と少なくとも50%の配列同一性を有することが望ましい。ハイブリダイゼーションプローブは、標識されたヌクレオチドの存在下でのPCR増幅、オリゴ標識化、ニックトランスレーション法、または末端標識化を利用して作製することができる。この核酸分子またはその断片を含むベクターを用いて、RNAポリメラーゼ及び標識したヌクレオチドを加えて*in vitro*でmRNAプローブを作製することができる。これらの方法はAPBが販売するキットを用いて行うことができる。

10

【0053】

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

20

30

【0054】

当分野で周知の方法でマイクロアレイを準備して分析することができる。オリゴヌクレオチドをマイクロアレイのプローブや標的として用いることができる。マイクロアレイを用いて、同時に極めて多数の遺伝子の発現レベルをモニタリングし、遺伝子変異体、突然変異及びSNP(一塩基多型)を同定することができる。このようなデータを用いて、遺伝子機能の解明や、症状及び疾患、または障害における遺伝子原理の解明や、症状及び疾患、障害の診断または治療、治療薬の開発、並びにこれらの治療薬の活性のモニタリングが可能である(例えば、Brennan他(1995)USPN 5,474,796; Schena他(1996)Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler他(1995)PCT出願WO95/251116; Shalon他(1995)PCT出願WO95/35505; Heller他(1997)Proc. Natl. Acad. Sci. 94:2150-2155; and Heller他(1997)USPN 5,605,662を参照)。

40

【0055】

50

ハイブリダイゼーションプローブはまた、天然のゲノム配列のマッピングに有用である。このプローブは、(1)特定の染色体、(2)染色体の特定の領域、(3)ヒト人工染色体(HACs)や酵母人工染色体(YACs)、細菌人工染色体(BACs)、細菌P1作製物、または単一の染色体、または(5)これらから作製されたcDNAライブラリなどの人工染色体作製物、にハイブリダイズすること可能である。

【0056】

発現

哺乳動物DSCR1L1タンパク質をコードする多数の核酸分子をベクターにクローニングして、このタンパク質若しくはその一部を宿主細胞で発現させることができる。この核酸配列を、DNAシャフリング(Stemmer and Cramer (1996) USPN 5,830,721)や部位特異的変異誘発などの方法によって、新規の制限部位を作り出したり、グリコシル化パターンを変えたり、優先コドンを変えて特定の宿主における発現を増大させたり、スプライスバリエーションを作り出したり、半減期を延長する等の操作が可能である。この発現ベクターは、特定の宿主における各要素の効率に基づいて選択された様々なサンプルに由来する転写及び翻訳調節エレメント(プロモーター及びエンハンサー、特定の開始シグナル、ポリアデニル化3'配列)を含み得る。in vitro組み換えDNA技術、合成技術及び/またはin vivo遺伝子組み換え技術を組み合わせて、このベクターに核酸配列と調節エレメントをつなぐことができる。このような技術は、当分野で周知であり、Sambrook(前出、ch. 4, 8, 16 and 17)に記載されている。

10

20

【0057】

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

30

【0058】

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

40

【0059】

長期に渡って組み換えタンパク質を産生させるために、同一或いは別のベクター上の選択マーカー遺伝子或いは可視マーカー遺伝子と共にこのベクターを持続的に細胞株に形質転換することができる。形質転換後、細胞を強化培地で約1~2日間増殖させてから選択培地に移す。選択マーカー、代謝拮抗物質、抗生物質、または除草剤耐性遺伝子は、関連する選択薬に対する抵抗性を与え、導入配列を確実に発現する細胞の増殖及び回収が可能となる。アントシアニン、緑色蛍光タンパク質(GFP)、グルクロニダーゼ、ルシフェラーゼなどの可視マーカーの発現によって同定された、或いは選択培地に生存することによって同定された耐性クローンを、培養技術を用いて増殖することができる。また、可視マーカーを用いて、導入された遺伝子によって発現するタンパク質を定量することができ

50

る。宿主細胞が目的の哺乳動物核酸分子を含むか否かの決定は、DNA-DNAまたはDNA-RNAハイブリダイゼーション、或いはPCR増幅技術に基づいて行うことができる。

【0060】

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

10

【0061】

細胞培地からのタンパク質の回収

精製を容易にするために、ベクターに導入する異種部分は、グルタチオンSトランスフェラーゼ (GST)、カルモジュリン結合ペプチド (CBP)、6-His、FLAG、MYC等を含む。GST及びCBP、6-Hisはそれぞれ、グルタチオン及びカルモジュリン、金属キレート樹脂が結合した市販のアフィニティマトリックスを用いて精製される。FLAG及びMYCは、市販のモノクローナル抗体及びポリクローナル抗体を用いて精製される。目的のタンパク質配列と異種部分との間にタンパク分解切断部位を設けて、生成の後の分離が容易にすることができる。組み換えタンパク質の発現及び精製の方法は Ausubel (前出、unit 16) に記載され市販されている。

20

【0062】

ペプチドの化学合成

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

30

40

【0063】

抗体の準備及びスクリーニング

ヤギ、ウサギ、ラット、マウス、及びヒト等を含む様々な宿主は、哺乳動物DSCR1L

50

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

【0064】

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

【0065】

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

【0066】

本哺乳動物DSCR1 タンパク質若しくはその一部を用いて、ファージミドまたはBリンパ球免疫グロブリン・ライブラリをスクリーニングして、目的の特異性を有する抗体を同定する。確立された特異性を有するモノクローナル抗体或いはポリクローナル抗体のいずれか一方を用いる、競合的結合またはイムノアッセイの様々なプロトコルが当分野で周知である。このようなイムノアッセイは通常、このタンパク質とその特異的な抗体との複合体形成の測定を行う。2つの非干渉エピトープに反応するモノクローナル抗体を用いる2部位モノクローナル系イムノアッセイが好ましいが、競合的結合によるアッセイを用いることもできる(Pound(1998) Immunochemical Protocols. Humana Press, Totowa NJ)。

【0067】

アッセイのための分子の標識化

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

【0068】

(診断)

本核酸分子、断片、オリゴヌクレオチド、相補的なRNA及びDNA分子、PNAを用いて、遺伝子発現の変化やmRNAの過剰な発現の不在/存在を検出及び定量、または治療

期間中の mRNA レベルのモニタリングを行うことができる。発現変動に関連する症状や疾患には、アルツハイマー病、ダウン症候群、ハンチントン病、及びピック病が含まれる。診断アッセイにハイブリダイゼーションまたは増幅技術を用いて、遺伝子発現の変化を検出するべく、患者からの生体サンプルの遺伝子発現レベルを標準的なサンプルの値と比較する。質的または量的なこのような比較法は当分野で周知である。

【0069】

例えば、本核酸分子またはプローブを標準的な方法で標識して、これをハイブリダイゼーション複合体の形成に好適な条件下で、患者からの生体サンプルに加える。インキュベーションの後、このサンプルを洗浄し、ハイブリダイゼーション複合体に関連する標識（シグナル）の量を定量して標準値と比較する。患者のサンプルにおける標識の量が標準値と著しく異なっている（高いまたは低い）場合は、関連する症状や疾患、または異常症の存在が示唆される。

10

【0070】

遺伝子発現に関連する症状や疾患、または異常症の診断のための基準を設けるために、正常或いは標準的な発現プロフィールを確立する。この発現プロフィールは、動物かヒトの正常な被験体から採取した生体サンプルを、ハイブリダイゼーションまたは増幅に好適な条件下で、プローブと結合させることによって確立することができる。標準的なハイブリダイゼーションは、正常な被験体から得た値と、実質的に精製された標的配列を所定量用いた実験値とを比較することによって定量することができる。このように求めた標準値を、特定の症状や疾患、または異常症を示す患者のサンプルから得た値と比較することができる。標準値と特定の症状に関連する値との偏差からその症状を診断する。

20

【0071】

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

【0072】

免疫学的方法

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

30

【0073】

（治療）

例えばRNA結合ドメインと保存されたセリン/プロリン・モチーフの文字列における化学的および構造的な類似性が、SEQ ID NO: 2の或る領域と、ヒトDSCR1、ヒト（g1435040）、及びヒトDSCR1L2（g6017919）などの他のDSCR1ファミリータンパク質の或る領域との間に存在する。発現または活性の上昇に関連する症状の治療においては、タンパク質の発現または活性を低下させることが望ましい。また、発現または活性の低下に関連する症状の治療においては、タンパク質の発現または活性を増大させることが望ましい。

40

【0074】

一実施例において、本哺乳動物タンパク質の発現または活性の変化に関連する症状の治療または予防のために、本哺乳動物タンパク質、その一部、またはその誘導体を患者に投与

50

することが可能である。このような疾患の例には、限定するものではないがアルツハイマー病、ダウン症候群、ハンチントン病、及びピック病が含まれる。

【0075】

別の実施例では、医薬用担体と共に精製された哺乳動物タンパク質を含む医薬組成物を患者に投与して、限定するものではないが上記した疾患を含むこの内在性タンパク質の活性、または発現の変化に関連する症状の治療または予防を行うことが可能である。

【0076】

更なる実施例では、この哺乳動物タンパク質の活性を調節するリガンドを患者に投与して、限定するものではないが上記した疾患を含むこのタンパク質の寿命や発現、または活性の変化に関連する症状の治療または予防を行うことが可能である。一実施態様では、この哺乳動物タンパク質に特異的に結合する抗体を、この哺乳動物タンパク質を発現する組織や細胞に薬剤を送達するためのターゲティング或いは送達機構として用いることもできる。

10

【0077】

更なる実施例では、この哺乳動物タンパク質またはその一部や誘導体を発現可能なベクターを患者に投与して、限定するものではないが上記した疾患を含むこのタンパク質の寿命や発現、または活性の変化に関連する症状の治療または予防を行うことが可能である。

【0078】

更なる実施例では、この核酸分子やその断片の相補配列を発現するベクターを患者に投与して、限定するものではないが上記した疾患を含むこのタンパク質の寿命や発現、または活性の変化に関連する症状の治療または予防を行うことが可能である。

20

【0079】

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

【0080】

核酸を用いる遺伝子発現の調節

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

30

40

【0081】

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

【0082】

50

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

10

【0083】

スクリーニングアッセイ

本哺乳動物タンパク質をコードする核酸分子を用いて、分子のライブラリをスクリーニングして特異的な結合親和性を調べることが可能である。このライブラリは、生物系において核酸分子の活性、複製、転写、または翻訳を調節するDNA分子、RNA分子、PNA、ペプチド、転写因子などのタンパク質、エンハンサー、リプレッサー、およびその他のリガンドを含み得る。このアッセイは、特異的な結合が許容される条件下で、哺乳動物核酸分子またはその断片を分子のライブラリと結合させるステップと、特異的な結合を検出して、その核酸分子と特異的に結合する少なくとも1つの分子を同定するステップとを含む。

20

【0084】

同様に、本哺乳動物タンパク質またはその一部を用いて、任意の様々なスクリーニングアッセイで分子または化合物のライブラリをスクリーニングすることができる。このようなスクリーニングに用いるタンパク質の一部は、溶液中に遊離している、非生物的若しくは生物学的な基板に固定されている、または細胞内に位置していてもよい。そのタンパク質と分子との特異的な結合を測定することが可能である。スクリーニングするライブラリの種類によって、アッセイでそのタンパク質に特異的に結合するDNA、RNA、PNA分子、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、ペプチド、タンパク質、薬剤、または他の任意のリガンドを同定することが可能である。極めて少ないアッセイ容量および極少量の試験化合物を用いるハイスループットのスクリーニング方法が、米国特許第5,876,946号に記載されている。この方法は、酵素阻害または受容体結合を調べるために多数の分子をスクリーニングする。

30

【0085】

リガンドの精製

本核酸分子またはその断片を用いてサンプルからリガンドを精製することが可能である。哺乳動物核酸分子またはその断片を用いてリガンドを精製する方法は、特異的な結合が許容される条件下でその核酸分子またはその断片とサンプルとを結合させるステップと、特異的な結合を検出するステップと、結合したタンパク質を回収するステップと、好適な試薬を用いて精製されたリガンドから核酸分子を分離するステップとを含む。

【0086】

同様に、本タンパク質またはその一部を用いてサンプルからリガンドを精製することが可能である。哺乳動物タンパク質またはその一部を用いてリガンドを精製する方法は、特異的な結合が許容される条件下でそのタンパク質またはその一部とサンプルとを結合させるステップと、そのタンパク質をリガンドとの間の特異的な結合を検出するステップと、結合したタンパク質を回収するステップと、好適なカオトロピック剤を用いて精製されたリガンドからタンパク質を分離するステップとを含む。

40

【0087】

薬理学

医薬組成物とは、所望の目的を達成するのに効果的な量の活性成分を含んでいる物質である。効果的な薬用量の決定は、当分野の技術者の能力による部分が大きい。どんな化合物

50

であっても、初めは細胞培養アッセイ或いは動物モデルの何れかによって治療効果のある薬用量を推定する。また、動物モデルを使って、好適な濃度範囲および投与経路を決定する。次に、このような情報を用いて、ヒトへの効果的な投与経路および薬用量を決定する。

【0088】

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

10

【0089】

モデル系

動物モデルを用いて、ヒトに相当する暴露条件で動物モデルがヒトに類似の毒物反応を示すバイオアッセイを行うことができる。哺乳動物が最も一般的なモデルである。大抵の毒物研究は、低コストで入手が容易であり、かつ参照できる中毒学が豊富であることから、ラットやマウスなどの齧歯類を用いて行われる。齧歯類近交系は、目的の遺伝子の過剰或いは過少な発現の生理学的な原因を調査するのに有用なモデルであり、疾患の診断および治療方法の開発にも有用である。特定の遺伝子を大量に発現する（例えば、乳汁中に分泌される）同系哺乳動物は、その遺伝子によって発現されるタンパク質の便利な供給源となり得る。

20

【0090】

中毒学

中毒学とは、生物系における物質の影響を研究する学問である。殆どの毒物研究はラットまたはマウスを用いて、物質がヒトの健康に与える影響を推定する。生理機能及び行動、恒常性プロセス、致死率における質的及び量的変化を観察して、毒性プロファイルを作成し、その物質に曝露された後のヒトの健康状態を評価する。

【0091】

遺伝子毒物学は、物質が遺伝子の突然変異を引き起こす能力を同定し分析する。遺伝毒性物質は通常、核酸との相互作用を促進する共通の化学的或いは物理的特性を有し、突然変異した染色体が子孫に受け継がれるのが最大の害である。毒物研究によって、受胎前の両親のどちらか一方、または妊娠中の母、発生段階の生物に投与された場合の、子孫における構造的或いは機能的な異常の頻度を増加させる物質を同定することが可能である。マウス及びラットは生殖周期が短く、統計的必要性を満たす多数の子を出産する能力から、これらの試験にはマウス及びラットが用いられる場合が最も多い。

30

【0092】

急性毒物の検査は、被験体への物質の一回の投与に基づき、その物質による症状または致死率を決定する。この検査では、（1）初めの投与量の範囲を決定する実験と、（2）有効な投与量の範囲を狭める実験と、（3）用量応答曲線を確立する実験の3つの実験が行われる。

40

【0093】

長期に渡る毒性検査では繰り返し物質を投与する。このような検査には、ラットやイヌが一般的に用いられ、分類学上異なった種からデータを収集する。物質を高い投与濃度で3~4ヶ月間、毎日投与することで、発癌を除く、成体動物における殆どの毒性の種類が明らかになるという研究結果が多数報告されている。

【0094】

一年或いはそれ以上の長期に渡る慢性毒性検査は、物質に毒性がないこと、或いは物質の発癌の可能性の何れかを実証するために行われる。ラットで検査が行われる場合、少なくとも3つの検査グループと1つの対照グループが用いられ、最初から最後まである間隔で

50

検査及びモニタリングが行われる。

【0095】

遺伝子組換え動物モデル

目的の遺伝子を過剰或いは過小に発現する遺伝子組換え齧歯類を同系交配し、それを用いてヒト疾患モデルを作製したり、治療薬検査や毒物検査を行う（例えば、米国特許第4,736,866号、同第5,175,383号、および同第5,767,337号を参照）。場合によっては、導入遺伝子が、胚発生中若しくは出生後の特定の時期に特定の種類の組織で活性化され得る。導入遺伝子の発現は、実験的薬剤治療を施す前、その最中、またはその後の、遺伝子組換え動物における表現型や組織特異的なmRNAの発現の分析からモニタリングすることができる。

10

【0096】

胚性幹細胞

齧歯類胚から単離された胚性幹細胞（ES細胞）は、胚を形成する可能性を維持している。ES細胞がキャリアとなる胚の中に導入されると、正常な発生が再開され、生まれる動物の全ての組織の一部を担うことになる。ES細胞は、実験用のノックアウトおよびノックイン齧歯類系を作製するのに好適な細胞である。マウス129/SvJ細胞株などのマウスES細胞は、マウスの初期胚から採取されてから当分野で周知の培養条件下で増殖されたものである。ノックアウト系に用いるベクターには、マーカー遺伝子配列を含むように破壊された疾患候補遺伝子が含まれる。このマーカー遺伝子配列は、in vivoでその転写および/または翻訳を阻害する。このベクターを、当分野で周知の電気穿孔法、

20

【0097】

また、in vitroでの神経細胞および造血系、心筋細胞などの様々な細胞型や組織の分化の研究に、ES細胞が用いられている（Bain 他（1995）Dev. Biol. 168:342-357; Wiles and Keller（1991）Development 111:259-267; and Klug 他（1996）J. Clin. Invest. 98:216-224）。近年の研究により、ヒト胚盤胞由来ES細胞をin vitroで操作して、内胚葉および中胚葉、外胚葉性の細胞型を含む8つの別の細胞系譜に分化可能であることが実証された（Thomsonら（1998）Science 282:1145-1147）。

30

【0098】

ノックアウト分析

遺伝子ノックアウト分析では、ヒト疾患遺伝子候補のある領域が、ネオマイシン・ホスホトランスフェラーゼ遺伝子などの非哺乳動物遺伝子を含むように酵素によって改変される（neo; Capecchi（1989）Science 244:1288-1292）。挿入されたコード配列は、標的遺伝子の転写および翻訳を阻害して疾患候補タンパク質の生化学的な合成を阻止する。この改変した遺伝子を培養胚性幹細胞（上記）に形質転換し、この形質転換細胞を齧歯類胚に注入し、この胚を偽妊娠メスに移植する。この遺伝子組換え子孫をクロス交配して、ホモ接合近交系を作り出す。

40

【0099】

ノックイン分析

胚発達の初期段階に現われる分化全能性ES細胞を用いてノックインヒト化動物（ブタ）または遺伝子組換えヒト疾患動物モデル（マウスまたはラット）を作り出すことができる。ノックイン技術を用いて、ヒト遺伝子のある領域を動物ES細胞に注入し、そのヒト配列が組換えによって動物細胞ゲノムの中に組み込まれるようにする。ヒト遺伝子が組み込

50

まれた全能性ES細胞を上記したように操作する。ヒトの症状に類似の情報を収集するべく、この近交系動物の研究および処置を行う。これらの方法を用いて、幾つかのヒト疾患モデルを作り出す(例えば、Lee 他(1998) Proc. Natl. Acad. Sci. 95:11371-11376; Baudoin 他(1998) Genes Dev. 12:1202-1216; and Zhuang 他(1998) Mol. Cell Biol. 18:3340-3349)。

【0100】

非ヒト霊長類モデル

動物実験の分野は、生理学および遺伝学、化学、薬理学、統計学等の基本科学から得たデータおよび方法論を利用する。このようなデータは、非ヒト霊長類がヒトの健康に関連し得るため、非ヒト霊長類における治療薬の効果の評価に重要である。ヒトの代わりにサルをワクチンおよび薬剤の評価に用いる。サルの反応は、類似の条件下でのヒトへの暴露に相当する。カニクイザル(Macaca fascicularis、Macaca mulatta)、および一般的なマーモセット(Callithrix jacchus)は、このような実験に用いられる最も一般的な非ヒト霊長類(NHP)である。NHPの群体作りおよび維持には多大な費用が掛かるため、初期の研究および毒物学的検査は、通常は齧歯類モデルで行われる。薬物嗜癖などの行動を調べる研究では、NHPは最良の実験動物である。更に、個々のNHPおよびヒトは、多くの薬剤および毒物に対して異なった感受性を示すため、これらの物質を「高代謝型」および「低代謝型」に分類することができる。

10

20

【0101】

更なる実施例では、限定するものではないが、トリプレット遺伝子コードおよび特異的な塩基対相互作用などの特性を含む、現在知られている核酸配列の特性に新しい技術が依存する場合は、本哺乳動物タンパク質をコードする核酸分子を、開発中のあらゆる分子生物学技術に用いることが可能である。

【0102】

(実施例)

本発明は、記載した特定の装置及び物質、方法に限定されるものではないことを理解されたい。特定の実施例について説明するが、同等の実施例を用いて本発明を具現することも可能である。本発明の範囲は、前記請求の範囲によってのみ限定されるものであって、記載した実施例によって限定されるものではない。また、以下に記載の実施例は、本発明を例示するためのものであって本発明を限定するものではない。例示目的で、ヒト海馬cDNAライブラリHIPONON02の作製方法を記載する。

30

【0103】

1 cDNAの作製

ヒト海馬cDNAライブラリHIPONON02は、頭蓋内出血で死亡した72歳の白人女性の海馬から採取した組織を用いて作製した。凍結組織を、POLYTRONホモジナイザー(PT-3000; Brinkmann Instruments, Westbury, NY)を用いて、TRIZOL試薬(1gの組織にTRIZOL 10ml; Life Technologies)においてホモジナイズし溶解した。ホモジナイズした後、クロロホルム(1:5 v/v クロロホルム:ホモジネート)を加え、溶解物を遠心分離した。水層を取り除いて、イソプロパノールでRNAを沈殿させた。RNAをDEPC処理水で再懸濁し、27℃で35分間、DNアーゼI(Life Technologies)で処理した。このRNAを酸性フェノールクロロホルム(pH4.7)で一回再抽出して、0.3M酢酸ナトリウムおよび2.5倍量のエタノールで沈殿させた。

40

【0104】

メッセンジャーRNA(mRNA)をOLIGOTEXキット(QIAGEN, Inc., Chatsworth, CA)を用いて単離した。これをcDNAライブラリの作製に用いた。このmRNAを、mRNAのポリ(A)尾部で一本鎖cDNAの合成を開始するように設計されたNotIプライマー-アダプターを含むSUPERSCRIPT

50

プラスミドシステム (Life Technologies) の推奨プロトコルに従って処理した。二本鎖 cDNA を平滑化し、EcoRI アダプターに結合し、NotI (New England Biolabs, Beverly MA) で消化した。この cDNA を、SEPHAROSE CL-2B カラム (APB) 上で分画化し、800 bp を越える大きさの cDNA を、PSPORT1 プラスミド (Life Technologies) の NotI 及び EcoRI 部位に結合させた。この組換えプラスミドを、DH5、DH12s、DH10B コンピテント細胞 (全て Life Technologies) に形質転換した。

【0105】

2 cDNA ライブラリの標準化

例を示すために、ヒト脳ライブラリの標準化について説明する。大腸菌株 DH12S コンピテント細胞 (Life Technologies) における HIPONOT01 プラスミドライブラリの約 1.13×10^6 の独立したクローンを、カルベニシリン (25 mg/l) 及びメチシリン (1 mg/ml) 選択下で、培地で成長させ、その後に電気穿孔法により形質転換を行った。ライブラリの存在レベルに従って余剰の cDNA 複製物の数を減らすために、cDNA ライブラリを以下の変更点を除き Soares 他 (1994, Proc Natl Acad Sci 91: 9228-9232) の方法に従って 1 回標準化した。変更点は次の通りである。プライマー伸長反応における鋳型に対するプライマーの比率を 2:1 から 10:1 に高めた。反応液の ddNTP 濃度を、より長いプライマー伸長物 (400 ~ 1000 塩基) が形成できるように各 ddNTP を 150 μ M まで低下させた。アニーリングハイブリダイゼーションを 13 時間から 48 時間に延長した。標準化したライブラリの一本鎖 DNA サークルを、ヒドロキシアパタイトクロマトグラフィーで精製し、ランダムプライミングにより部分的な二本鎖に変換し、大腸菌株 DH10B コンピテント細胞 (Life Technologies) の中に電気穿孔法により導入した。

【0106】

3 pINCY プラスミドの作製

このプラスミドは、pSPORT1 プラスミド (Life Technologies) を EcoRI 制限酵素 (New England Biolabs, Beverly MA) で消化し、オーバーハングした端部を、クレノウ酵素 (New England Biolabs) 及び 2'-デオキシヌクレオチド 5'-三リン酸 (dNTPs) を用いて二本鎖に合成した。このプラスミドが自己連結した後、このプラスミドで細菌宿主となる大腸菌株 JM109 を形質転換した。

【0107】

細菌 (pSPORT1-RI) によって生成された中間プラスミドは、EcoRI では消化されず、HindIII (New England Biolabs) で消化し、オーバーハングした端部をクレノウ酵素及び dNTPs で二本鎖にした。リンカー配列をリン酸化し、5' 平滑末端に繋ぎ、EcoRI で消化し、自己連結させた。JM109 宿主細胞を形質転換した後、プラスミドを単離し、HindIII ではなく EcoRI で選択的な消化の検査を行った。この基準を満たす 1 つのコロニーを pINCY プラスミドと呼ぶ。

【0108】

NotI 及び EcoRI 制限酵素を用いて作製したライブラリから cDNA を取り込むプラスミドの能力を検査した後、いくつかのクローンをシーケンシングし、約 0.8 kb のインサートを含む 1 つのクローンを選択し、そのクローンから大量のプラスミドを生成した。NotI 及び EcoRI で消化した後、ライブラリの作製に用いるためにこのプラスミドをアガロースゲル上で分画し、QIAQUICK カラム (Qiagen) を用いて精製した。

【0109】

4 cDNA クローンの単離及びシーケンシング

10

20

30

40

50

プラスミドDNAを細胞から遊離し、MINIPREPキット(Edge Biosystems, Gaithersburg MD)またはREAL Prep 96プラスミドキット(Qiagen)のいずれかを用いて精製した。このキットには、96ウェルブロック、及び960回の精製を行うための試薬が含まれている。以下の変更点を除き、推奨プロトコルに従った。(1)細菌を、25mg/lのカルベニシリン及び2.4%のグリセロールと共に1mlの滅菌TERRIFIC BROTH(BD Biosciences, San Jose CA)において培養した。(2)19時間インキュベートしてから、0.3mlの溶解バッファで溶解した。(3)イソプロパノール沈殿の後、プラスミドDNAのペレットを0.1mlの蒸留水で再懸濁した。プロトコルの最終ステップの後、サンプルを96ウェルブロックに移し、4 で保管した。

10

【0110】

MICROLAB 2200システム(Hamilton, Reno NV)及びDNA ENGINEサーマルサイクラー(MJ Research)を用いて、シークエンシングするためにcDNAを調整した。このcDNAを、ABI PRISM 377シークエンシングシステム(Applied Biosystems)またはMEGABASE 1000 DNAシークエンシングシステム(APB)を用いてSanger and Coulson(1975; J Mol Biol 94: 441-448)の方法でシークエンシングした。単離したもののほとんどを、標準的なABIプロトコル及びキット(Applied Biosystems)に従って、0.25x~1.0xの濃度の溶液容量でシークエンシングした。別法では、APBから入手した溶液及び色素を用いてcDNAをシークエンシングした。

20

【0111】

5 cDNA配列の伸長

本cDNAは、cDNAクローン及びオリゴヌクレオチドプライマーを用いて伸長した。一方のプライマーは既知の断片の5'の伸長を開始するために合成し、他方のプライマーは既知の断片の3'の伸長を開始するために合成した。開始プライマーは、OLIGプライマー分析ソフトウェア(National Biosciences Insights)を用いて、約22~約30個のヌクレオチドの長さ、約50%以上のGC含量で、685~72 の温度で標的配列にアニールするように設計した。ヘアピン構造及びプライマー-プライマー二量体を形成するヌクレオチドのストレッチは排除した。

30

【0112】

選択されたcDNAライブラリを鋳型として用いてこの配列を伸長した。2段階以上の伸長が必要な場合は、追加の或いは入れ子状のプライマー(nested primer)を設計した。好適なライブラリは、大きなcDNAを含むように大きさが選択されたライブラリである。ゲノムライブラリは、調節エレメントを得るためにプロモーター結合領域の5'を伸長する際に有用である。

【0113】

米国特許第5,932,451号に開示されているような方法を用いてPCR法で高い忠実度の増幅を達成した。PCRは、DNA ENGINEサーマルサイクラー(MJ Research)を用いて96ウェルプレートで行った。反応混合液には、鋳型DNA及び200nmolの各プライマー、反応緩衝液(Mg²⁺、(NH₄)₂SO₄、β-メルカプトエタノールを含む)、Taq DNAポリメラーゼ(APB)、ELONGASE酵素(Life Technologies)、Pfu DNAポリメラーゼ(Stratagene)が含まれている。プライマーの組、PCI A及びPCI B(Incyte Pharmaceuticals)に対して以下のパラメーターで増幅を行った。

40

ステップ1	94	で3分間
ステップ2	94	で15秒
ステップ3	60	で1分間
ステップ4	68	で2分間
ステップ5	ステップ2、3、及び4を20回繰り返す	

50

ステップ 6 68 で5分間

ステップ 7 4 で保管。

別法では、プライマーの組、T7とSK+(Stratagene)に対して以下のパラメーターで増幅を行った。

ステップ 1 94 で3分間

ステップ 2 94 で15秒

ステップ 3 57 で1分間

ステップ 4 68 で2分間

ステップ 5 ステップ 2、3、及び 4 を 20 回繰り返す

ステップ 6 68 で5分間

ステップ 7 4 で保管。

10

【0114】

各ウェルのDNA濃度は、100 μ lのPICOGREEN定量試薬(1 \times TEにおける試薬0.25%(v/v); Molecular Probes)及び0.5 μ lの希釈していないPCR産物を不透明な蛍光光度計プレート(Corning, Acton MA)の各ウェルに分注して、DNAがその試薬と結合できるようにして測定した。このプレートをFluoroskan II(Labsystems Oy)でスキャンして、サンプルの蛍光を測定してDNAの濃度を定量化した。反応混合物の5~10 μ lのアリコットを1%のアガロースミニゲル上で電気泳動して解析し、何れの反応がより長い配列の伸長に成功したかを決定した。

20

【0115】

伸長したクローンを脱塩及び濃縮してから384ウェルプレートに移し、CviJIコラウシルスエンドヌクレアーゼ(Molecular Biology Research, Madison WI)で消化し、pUC18ベクター(APB)に再連結する前に音波処理または切断した。ショットガンシーケンシングのために、消化したヌクレオチド配列を低濃度(0.6%~0.8%)のアガロースゲル上に分離させ、断片を切断し、ゲルをAGARACE酵素(Promega)で消化した。T4 DNAリガーゼ(New England Biolabs)を用いて伸長したクローンをpUC18ベクター(APB)に再連結し、Pfu DNAポリメラーゼ(Stratagene)で制限部位の伸び出しを処理してから、大腸菌コンピテント細胞に形質転換した。形質転換細胞が抗生物質を含む培地で選択され、それぞれのコロニーを切りとって、LB/2Xカルベニシリン培養液の入った384ウェルプレートの中で、37 で一晩培養した。

30

【0116】

この細胞を溶解して、Taq DNAポリメラーゼ(APB)及びPfu DNAポリメラーゼ(Stratagene)を用いて以下の手順でDNAを増幅した。

ステップ 1 94 で3分間

ステップ 2 94 で15秒

ステップ 3 60 で1分間

ステップ 4 72 で2分間

ステップ 5 ステップ 2、3、及び 4 を 29 回繰り返す

ステップ 6 72 で5分間

ステップ 7 4 で保管。

40

上記したようにPICOGREEN試薬(Molecular Probes)でDNAを定量した。DNA回収率の悪いサンプルは、上記した条件で再び増幅した。サンプルを20%のジメチルサルホサイド(dimethylsulphoxide)(1:2, v/v)で希釈し、DYENAMICエネルギー移動シーケンシングプライマー及びDYENAMIC DIRECTサイクルシーケンシングキット(APB)またはABI PRISM BIGDYE Terminator cycle sequencing ready reactionキット(Applied Biosystems)を用いてシーケンシングした。

50

【0117】

6 cDNA及びそれらから導き出されたタンパク質の相同性検索

配列表のcDNA及びそれらから導き出されたアミノ酸配列を、GenBank、SwissProt、BLOCKSなどのデータベースにおいて検索する。すでに同定されアノテーションの付けられた配列またはドメインを含むこれらのデータベースを、BLASTまたはBLAST2(Altschulら、前出; Altschul、前出)を用いて検索し、アラインメントを作成し、完全に一致する或いは相同な配列を決定する。このアラインメントは、原核生物(細菌)または真核生物(動物、真菌、または植物)を起源とする配列に対するものである。或いは、Smith and Smith(1992, Protein Engineering 5: 35-51)に記載されているようなアルゴリズムを用いて、一次配列パターン及び二次構造ギャップペナルティを処理することが可能である。本明細書に記載した全ての配列は、その長さが少なくとも49ヌクレオチドであり、不要な塩基(A、C、G、或いはTではなくNと記載)は12%以下である。

10

【0118】

Karlin(前出)に詳述されているように、問い合わせ配列とデータベース配列との間のBLASTによる一致を統計学的に評価し、ヌクレオチドに対しては $10^{-2.5}$ 、ペプチドに対しては $10^{-1.4}$ の閾値を満たす一致のみを記録した。相同性はまた、以下のように計算した積スコアによって評価した。まず、BLASTにおける核酸またはアミノ酸同一性(問い合わせ配列と参照配列との間の)のパーセントに、最大可能BLASTスコアのパーセント(問い合わせ配列及び参照配列の長さに基づく)を乗じてから、100で除した。実験室で用いられるハイブリダイゼーション法と比べ、完全一致に対する電子的なストリンジェンシーを70と設定し、完全一致に対する保守的な下限を約40に設定した(不要な塩基対による1~2%のエラーを有する)。

20

【0119】

無料で利用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>)は、既知の核酸分子を整列するために用いる“blastn”、及び核酸分子或いはアミノ酸分子のいずれかの直接ペアワイズ比較に用いられるBLAST2を含む様々な配列分析プログラムを含む。BLASTプログラムは、一般に、例えば以下のようにデフォルト設定されたギャップ及びその他のパラメーターで実行される。

30

【0120】

Matrix: BLOSUM62
 Reward for match: 1
 Penalty for mismatch: -2
 Open Gap: 5 及び Extension Gap: 2 penalties
 Gap x drop-off: 50
 Expect: 10
 Word Size: 11
 Filter: on

40

同一性を、配列の全長或いは配列の一部に対して測定することができる。Brennerらが(1998; Proc Natl Acad Sci 95: 6073-6078、言及することをもって本明細書の一部とする)、配列同一性により構造の相同性を同定するBLASTの能力を分析した。この分析により、少なくとも150残基の配列アラインメントに対して30%の同一性が信頼できる閾値であり、少なくとも70残基の配列アラインメントに対して40%の同一性が信頼できる閾値である。

【0121】

本明細書の哺乳動物核酸分子を、LIFESEQ GOLDデータベースで見出した構築されたコンセンサス配列または鋳型と比較した。cDNA、伸長、完全長、及びショット

50

ガンシークエンシングプロジェクトに由来するコンポーネント配列を P H R E D で分析し、質のスコアを割り当てた。許容できる質スコアを有する全ての配列に対して様々なプリプロセッシング及び編集を行い、質の低い 3' 末端、ベクター及びリンカー配列、ポリ A 尾部、A 1 u リピート、ミトコンドリア及びリボソーム配列、及び細菌汚染配列を除去した。編集した配列は少なくともその長さが 5 0 b p でなければならない。情報の少ない配列、並びにジヌクレオチドリピート、A 1 u リピートなどの繰り返しエレメントは N で置換するかマスクした。

【 0 1 2 2 】

編集した配列を、配列が遺伝子ピンに割り当てられる構築処理を行った。それぞれの配列は 1 つのピンにのみ属し、鋳型を形成するべく各ピンにおける配列を構築した。新規にシーケンシングしたコンポーネントを B L A S T 及び C R O S S M A T C H を用いて存在するピンに加えた。ピンに加えるためにはこのコンポーネント配列は、B L A S T 質スコアが 1 5 0 以上であって、少なくとも 8 2 % の局所的な同一性のアラインメントでなければならない。それぞれのピンにおける配列を、P H R A P を用いて構築した。いくつかの重複するコンポーネント配列を有するピンは D E E P P H R A P を用いて構築した。それぞれの鋳型の向きは、そのコンポーネント配列の数及び向きに基づいて決定した。

10

【 0 1 2 3 】

それぞれのピンを互いに比較して、局所類似性が 8 2 % 以上のピンを 1 つにまとめて再構築した。局所同一性が 9 5 % 未満の鋳型を有するピンは分けた。鋳型は、S T I T C H E R / E X O N M A P P E R アルゴリズムで分析した。この S T I T C H E R / E X O N M A P P E R アルゴリズムは、スプライスバリエーション、択一的スプライシングエキソン、スプライスジャンクション、並びに組織の種類或いは疾患の状態により発現が異なる択一的にスプライシングされた遺伝子などの存在の確率を分析することができる。構築処理を繰り返し行い、G B p r i などの G e n B a n k データベースに対して B L A S T を用いて鋳型にアノテーションを付けた。完全一致は、2 0 0 塩基対に対する 9 5 % 以上の局所同一性から 1 0 0 塩基対に対する 1 0 0 % の局所同一性を有すると定義し、相同一致は、 1×10^{-8} の E 値 (確率スコア) を有すると定義した。また鋳型を、G E N P E P T に対してフレームシフト F A S T x 分析を行い、相同一致を 1×10^{-8} の E 値を有すると定義した。鋳型の分析及び構築については、1 9 9 9 年 3 月 2 5 日に出願された米国特許出願第 0 9 / 2 7 6 , 5 3 4 号に開示されている。

20

30

【 0 1 2 4 】

構築の後に、鋳型を B L A S T 、モチーフ、及びその他の機能分析法で分析し、1 9 9 7 年 3 月 6 日に出願の米国特許出願第 0 8 / 8 1 2 , 2 9 0 号及び同第 0 8 / 8 1 1 , 7 5 8 号、1 9 9 7 年 1 0 月 9 日に出願の米国特許出願第 0 8 / 9 4 7 , 8 4 5 号、及び 1 9 9 8 年 3 月 4 日に出願の米国特許出願第 0 9 / 0 3 4 , 8 0 7 号に記載されている方法を用いてタンパク質を階層に分類した。次に鋳型を、3 つ全ての前方読み枠にそれぞれの鋳型を翻訳して、それぞれの翻訳を、H M M E R ソフトウェアパッケージ (W a s h i n g t o n U n i v e r s i t y S c h o o l o f M e d i c i n e , S t . L o u i s M O ; <http://pfam.wustl.edu/>) を用いて隠れマルコフモデルをベースにしたタンパク質ファミリー及びドメインの P F A M データベースにおいて検索した。

40

【 0 1 2 5 】

本核酸分子を、M A C D N A S I S P R O ソフトウェア (H i t a c h i S o f t w a r e E n g i n e e r i n g) 及び L A S E R G E N E ソフトウェア (D N A S T A R) を用いて更に分析し、G e n B a n k の 齧 歯 類、哺乳類、脊椎動物、原核生物、及び真核生物のデータベース、S w i s s P r o t、B L O C K S、P R I N T S、P F A M、及び P r o s i t e などの公共のデータベースに対して問い合わせた。

【 0 1 2 6 】

7 染色体マッピング

S t a n f o r d H u m a n G e n o m e C e n t e r (S H G C)、W h i t e h

50

ead Institute for Genome Research (WIGR)、及び Genethon などの公共の情報源から入手可能な放射線ハイブリッド及び遺伝子マッピングデータを用いて、マッピングできる配列表の cDNA が存在するかを調べた。DSCR1 をコードする核酸分子のマッピングされた全ての断片が、同じ位置にマッピングされた全ての関連する調節及びコード配列に割り当てられた。遺伝子マップ位置は、ヒト染色体の範囲すなわち区間として表される。センチモルガン (cM) (ヒト DNA の 100 万塩基対に概ね相当する) で表されるマップ位置の範囲は、染色体 p 腕の末端から測定する。

【0127】

7 ハイブリダイゼーション技術及び分析

10

核酸分子の基板への固定

本 cDNA を以下に示す方法で基板に固定する。cDNA の混合液をジェル電気泳動により分画後、毛細管輸送によってナイロン膜に移す。別法では、cDNA を個別にベクターに結合して、それを細菌宿主細胞に挿入してライブラリを形成する。次に cDNA を以下の方法で基板に配列する。第 1 の方法では、個別のクローンを含む細菌細胞を機械的に摘み上げてナイロン膜に整列させる。このナイロン膜を、選択薬 (用いるベクターによって異なるが、カルペンシリン、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など) を含む LB 寒天培地に載置し、37 で 16 時間インキュベートする。この膜を寒天培地から取り除き、次にこの膜をコロニー側を上にして 10% SDS、変性溶液 (1.5 M NaCl、0.5 M NaOH)、中和液 (1.5 M NaCl、1 M Tris、pH 8.0) に入れ、2x SSC において 10 分間づつ 2 回インキュベートする。次にこの膜を、STRATALINKER UV クロスリンカー (Stratagene) で UV 照射する。

20

【0128】

第 2 の方法では、インサートに隣接したベクター配列に相補的なプライマーを用いて、PCR を 30 サイクル行い細菌ベクターから cDNA を増幅する。PCR 増幅により、拡散の濃度を開始時の 1~2 ng から最終的に 5 µg まで増大させる。約 400 bp ~ 約 5000 bp の増幅した拡散を SEPHACRYL-400 ピーズ (APB) を用いて精製する。精製した拡散を、手動で或いはドット/スロットプロットングマニフォールド及び吸入装置でナイロン膜に配列し、上記した変性、中和、及び UV 照射によって固定する。精製した拡散を、米国特許第 5,807,522 号に開示された方法でポリマーコートスライドガラスに機械的に配列して固定する。ポリマーコートスライドガラスは、顕微鏡スライドガラス (Corning) を 0.1% SDS 及びアセトンで超音波洗浄した後、4% フッ化水素酸 (VWR Scientific Products, West Chester PA) においてエッチングし、次に 95% エタノールにおいて 0.05% アミノプロピルシラン (Sigma-Aldrich, St Louis MO) でコーティングし、オープンで 110 で硬化させて準備する。処理中及び処理後にこのスライドガラスを蒸留水で十分に洗浄する。次に、核酸分子をスライド上に配列し、STRATALINKER UV クロスリンカー (Stratagene) を用いてこの核酸分子のアレイを UV 照射してスライド上に固定する。次にこのアレイを、室温で 0.2% SDS で洗浄し、蒸留水で 3 回すすぐ。非特異的な接合部位を、アレイをリン酸バッファー (Tropix, Bedford MA) に於ける 0.2% カゼインにおいて 60 で 30 分間インキュベートしてブロックし、次にこのアレイを 0.2% SDS で洗浄し、前記したように蒸留水で洗い流す。

30

40

【0129】

メンブレンを用いるハイブリダイゼーションのためのプローブの準備

本配列表の cDNA に由来するハイブリダイゼーションプローブは、メンブレンハイブリダイゼーションにおいて cDNA、mRNA、またはゲノム DNA をスクリーニングするために用いられる。プローブを準備するために、cDNA を 45 µl TE バッファにおいて 40~50 ng の濃度に希釈し、100 で 5 分間加熱して変性し、短時間遠心分

50

離する。次に、変性した cDNA を REDIPRIME チューブ (APB) に加え、青色が十分に拡散するまで軽く混合し、短時間遠心分離する。5 μ l の [³²P] dCTP をチューブに加え、その内容物を 37 で 10 分間インキュベートする。この標識化反応を 5 μ l の 0.2 M EDTA を加えてストップし、PROBEQUANT G-50 マイクロカラム (APB) を用いてプローブを組み込まれなかったヌクレオチドから精製する。精製したプローブを 100 で 5 分間加熱してから氷上で 2 分間冷却し、以下に記載するようにメンブレンハイブリダイゼーションに用いる。

【0130】

ポリマーコートスライドを用いるハイブリダイゼーションのためのプローブの準備

サンプルから単離した mRNA に由来するハイブリダイゼーションプローブを用いて、アレイハイブリダイゼーションにおいて本配列表の cDNA をスクリーニングする。プローブは、GEMbright キット (Incyte Genomics) を用いて、mRNA を 9 μ l TE バッファにおいて 200 ng の濃度に希釈し、5 μ l の 5x バッファ、1 μ l の 0.1 M DTT、3 μ l の Cy3 または Cy5 標識混合物、1 μ l の RNアーゼインヒビター、1 μ l の逆転写酵素、及び 5 μ l の 1x 酵母コントロール mRNA を加えて準備する。酵母コントロール mRNA は、非コード酵母ゲノム DNA (W. Lei, unpublished) から *in vitro* 転写によって合成する。量的コントロールとして、0.002 ng、0.02 ng、0.2 ng、及び 2 ng のコントロール mRNA からなるセットを、サンプル mRNA に対してそれぞれ 1:100, 000、1:10, 000、1:1000、及び 1:100 (w/w) の比率で逆転写反応液に希釈する。mRNA の異なった発現パターンを検査するために、第 2 のセットのコントロール mRNA を、1:3、3:1、1:10、10:1、1:25、及び 25:1 (w/w) の比率で逆転写反応液に希釈する。反応液を混合し、37 で 2 時間インキュベートする。次に反応液を 85 で 20 分間インキュベートし、2 つの連続した CHROMA SPIN+TE 30 カラム (Clontech) を用いてプローブを精製する。精製したプローブを、DEPC 処理水で 90 μ l に希釈して、2 μ l の 1 mg/ml グリコーゲン、60 μ l の 5 M 酢酸ナトリウム、300 μ l の 100% エタノールを加えてエタノール沈殿させる。このプローブを 20, 800 x g において 20 分間遠心分離し、ペレットを 12 μ l の再懸濁バッファに再懸濁させ、65 で 5 分間加熱し、完全に混合する。上記したようにプローブを加熱及び混合してから氷上で保管する。プローブを以下に記載するように高密度のアレイハイブリダイゼーションに用いる。

【0131】

メンブレンを用いるハイブリダイゼーション

メンブレンを、1% Sarkosyl 及び 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において 55 で 2 時間プレハイブリダイゼーションする。15 ml の新しいハイブリダイゼーション溶液に希釈したプローブをメンブレンに加える。55 で 16 時間、メンブレンにプローブをハイブリダイズさせる。ハイブリダイゼーションの後、メンブレンを 1 mM Tris (pH 8.0)、1% Sarkosyl において 25 で 15 分間洗浄し、1 mM Tris (pH 8.0) において 25 で 15 分間づつ 4 回洗浄する。ハイブリダイゼーション複合体を検出するために、XOMAT-AR フィルム (Eastman Kodak, Rochester NY) をメンブレンに一晚 -70 で露出し、現像して目で確認する。

【0132】

ポリマーコートスライドを用いるハイブリダイゼーション

プローブを 65 で 5 分間加熱し、5415C マイクロ遠心分離器 (Eppendorf Scientific, Westbury NY) を用いて毎分 9400 回転で 5 分間遠心分離し、アレイの表面に 18 μ l のアリコットを載せカバースリップで覆う。このアレイを、顕微鏡スライドより僅かに大きいキャビティを有する防水容器に移す。この容器の角から 140 μ l の 5x SSC を加えてその内部を 100% の湿度に保つ。アレイを含

む容器を60℃で約6時間半インキュベートする。このアレイを、1×SSC、0.1% SDSにおいて45℃で10分間洗浄し、0.1×SSCにおいて45℃で10分間づつ3回洗浄した後、乾燥させる。

【0133】

ハイブリダイゼーション反応を、絶対ハイブリダイゼーション (absolute hybridization) 方式またはディファレンシャルハイブリダイゼーション方式で行う。絶対ハイブリダイゼーション方式の場合、1つのサンプルからのプローブをアレイの要素にハイブリダイズさせ、ハイブリダイゼーション複合体が形成された後にシグナルを検出する。シグナルの強度がサンプルにおけるプローブmRNAのレベルに相関する。ディファレンシャルハイブリダイゼーション方式の場合、2つの生体サンプルにおける遺伝子のセットの異なった発現を分析する。2つのサンプルからのプローブを準備し、異なった標識成分で標識する。2つの標識したプローブの混合物をアレイ要素にハイブリダイズさせ、2つの異なった標識からの蛍光シグナルをそれぞれ別に検出できる条件下でシグナルを検査する。アレイ上の要素に、2つの生体サンプルに由来するプローブのそれぞれが同数ハイブリダイズする場合、明瞭な複合蛍光が得られる (Shalton W095 / 35505)。

10

【0134】

ハイブリダイゼーション複合体をInnova 70混合ガス10Wレーザーを備えた顕微鏡 (Coherent, Santa Clara CA) で検出する。この顕微鏡は、Cy3を励起するために488nmのスペクトル線が発生し、Cy5を励起するために632nmのスペクトル線が発生することができる。励起レーザー光を20倍の顕微鏡対物レンズ (Nikon, Melville NY) を用いてアレイに集束させる。アレイを含むスライドを顕微鏡上のコンピュータ制御されたXYステージ上に載せ、20μmの解像度で対物レンズを通してラスタースキャンする。ディファレンシャルハイブリダイゼーション方式の場合、2つの蛍光物質を順番にレーザーで励起する。波長に応じて放出された光を2つの蛍光物質に対応する2つの光電子増倍管検出器 (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) に分割する。アレイと光電子増倍管との間に配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リングする。用いられる蛍光物質の最大発光波長は、Cy3が565nmであり、Cy5が650nmである。スキャンの感度は、プローブ混合物に加えられた酵母コントロールmRNAによって生成されるシグナル強度を用いて校正する。アレイのある位置に相補的なDNA配列が含まれていると、その位置におけるシグナルの強度が、ハイブリダイズ種 (hybridizing species) の重量比で1:100,000の関係となり得る。

20

30

【0135】

光電子増倍管の出力をIBM互換性パーソナルコンピュータにインストールした12ビットRTI-835Hアナログ/デジタル変換ボードを用いてデジタル化する (Analog Devices, Norwood MA)。このデジタル化したデータを、イメージとして表示する。この場合、シグナル強度を、20色線形変換を用いて、青 (弱いシグナル) から赤 (強いシグナル) の疑似色スケールにマッピングする。データはまた量的にも分析する。2つの異なった蛍光物質が同時に励起されて測定された場合、データを、各蛍光物質に対する発光スペクトルを用いて蛍光物質間の光学クロストーク (発光スペクトルの重複による) をまず補正する。各スポットからのシグナルがグリッドの各要素の中心に来るように、蛍光物質シグナル強度の上にグリッドを重ね合わせる。各要素内の蛍光シグナルを統合して、シグナルの平均強度に対応する数値を求める。シグナル分析に用いるソフトウェアはGEMTOOLSプログラム (Incyte Pharmaceutical) である。

40

【0136】

8 電子分析

BLASTを用いて、GenBankやLifeSeqデータベース (Incyte G

50

e n o m i c s) において同一または関連分子を検索した。ヒトおよびラット配列の積スコアを下記のようにして計算した。BLASTスコアにヌクレオチド一致率(%)を乗じ、その積を2つの配列の短い方の長さ×5で除し、短い配列の長さに対する100%のアラインメントが積スコア100とする。積スコアは2つの配列間の類似性の程度と配列一致の長さの両方を考慮する。例えば、積スコア40では一致はエラーが1%~2%以内の厳密さであり、積スコア70以上では一致は完全となる。類似分子または関連分子は通常、8~40の間の積スコアを示す分子を選択することで特定できる。

【0137】

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

10

【0138】

9 相補的な分子

約5(PNA)から5000bp(cDNAインサートの相補配列)の本核酸分子に相補的な分子を用いて、遺伝子発現の検出或いは阻害を行う。これらの分子をOLIGOプライマー分析ソフトウェア(Molecular Biology Insights)で選択した。検出については、実施例7に記載されている。プロモーターが結合するのを妨げて転写を阻害するために、相補的な分子を最もユニークな5'配列に結合すると共に読み枠の開始コドンの5'UTR上流のヌクレオチドを含むよう設計する。相補的な分子は、エンハンサーやイントロン等のゲノム配列を含み、この相補的な分子を用いて「三重らせん」塩基対を形成して、二重らせんが十分に開かないようにし、これによりポリメラーゼ、転写因子、または調節分子が結合できないようにする。翻訳を阻害するためには、リボソームが本哺乳動物タンパク質をコードするmRNAに結合するのを阻害するように相補的な分子を設計する。

20

【0139】

相補的な分子を発現ベクターに導入して、これを用いて細胞系を形質転換して効果を検査する。例えば、器官、腫瘍、滑腔、また脈管系を形質転換して一過性或いは短期治療の効果、また幹細胞、接合体、または他の再生システムを形質転換して長期或いは安定した遺伝子治療の効果を検査することが可能である。一過性発現は、非複製ベクターを用いると1ヶ月以上続き、ベクター複製を誘導する適切なエレメントを形質転換/発現系に用いると3ヶ月以上続く。

30

【0140】

相補的な分子をコードするベクターで好適な分裂細胞を安定的に形質転換することにより、遺伝子組換え細胞系、組織、または生物を作製することができる(USPN 4,736,866)。同化して十分な量のベクターを複製する細胞を安定的に組込むことが可能であるため、十分な相補的分子が産生され、本哺乳動物タンパク質をコードする核酸分子の活性を低下或いは完全に消失させることができる。

40

【0141】

10 DSCR1K1 の発現

本哺乳動物タンパク質の発現及び精製は、哺乳動物細胞発現系若しくは昆虫細胞発現系を用いて行うことができる。pUB6/V5-Hisベクター系(Invitrogen, Carlsbad CA)を用いて、CHO細胞でDSCR1K1を発現させる。ベクターは選択可能なbsd遺伝子、複数のクロニング部位、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。形質転換された細胞を、ブラストサイジン

50

を含む培地で選択する。

【0142】

Spodoptera frugiperda (Sf9) 昆虫細胞を組換え Autographa californica 核多角体病ウイルス (バキュロウイルス) で感染させる。この多角体遺伝子を、相同組換えにより哺乳動物核酸分で置換し、多角体プロモーターにより転写が促進されるようにする。このタンパク質を、上述したように精製を可能にする 6xhis で合成して融合タンパク質にする。精製されたタンパク質を次のように用いて、抗体を作製する。

【0143】

1.1 抗体の生産

DSCR1K1 をポリアクリルアミドゲル電気泳動法で精製して、マウスやウサギの免疫に用いる。抗体の生産は下記のプロトコルに従う。別法では、DSCR1K1 のアミノ酸配列を LASERGENE ソフトウェア (DNASTAR) で解析して免疫原性の高い領域を決定する。通常は C-末端近傍或いは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、抗体を産生させるために用いる。通常は、長さ約 15 残基のエピトープを、Fmoc ケミストリを用いる ABI 431A ペプチドシンセサイザ (Applied Biosystems) で作製し、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と反応させて K L H (Sigma-Aldrich) に結合させ、免疫原性を高める。

【0144】

完全フロイントアジュバントでエピトープ-K L H 複合体を用いてウサギを免疫する。次に、不完全フロイントアジュバントで間隔を置いて免疫化を繰り返す。マウスの場合には最低 7 週間、ウサギの場合には最低 12 週間が経過した後、抗血清を抽出して抗ペプチド活性を検査した。この検査では、ペプチドをプラスチックに結合させ、1% のウシ血清アルブミンでブロックし、ウサギ抗血清で反応させ、洗浄し、放射性ヨウ素化標識ヤギ抗ウサギ IgG で反応させる。当分野で公知の方法を用いて、抗体価と形成された複合体の量を決定する。

【0145】

1.2 特異的抗体を用いる天然タンパク質の精製

天然或いは組換えタンパク質を、そのタンパク質と特異的に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て作製される。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、そのタンパク質の優先的な吸着を可能にする界面活性剤の存在下で、高イオン強度バッファーでカラムを洗浄する。そのタンパク質が結合した後に、pH 2~3 或いはそれ以上のバッファー、高濃度の尿素、またはチオシネートイオンを用いて抗体とタンパク質との結合を切断してカラムから溶出させ、そのタンパク質を回収する。

【0146】

1.3 核酸分子またはタンパク質と特異的に結合する分子のスクリーニング

本核酸分子またはそのタンパク質をそれぞれ、³²P-dCTP、Cy3-dCTP、または Cy5-dCTP (APB) と、BIODIPY または FITC (Molecular Probes, Eugene OR) とで標識する。前もって基質上に配列した候補分子または複合体のライブラリを、標識した核酸分子またはタンパク質の存在下でインキュベートする。核酸配列或いはアミノ酸配列に適した条件下でインキュベートした後、基板を洗浄し、特異的な結合即ち複合体成形を示す標識を保持した基板上の全ての位置をアッセイしてリガンドを同定する。様々な濃度の核酸またはタンパク質から得たデータを使用して、標識した核酸或いはタンパク質と結合した分子との間の親和性を計算する。

【0147】

1.4 2-ハイブリッドスクリーン

酵母 2-ハイブリッド法、即ち MATCHMAKER LexA Two-Hybrid シ

10

20

30

40

50

ステム (Clontech Laboratories, Palo Alto CA) を用いて、本発明の哺乳動物タンパク質と結合するペプチドをスクリーニングする。本タンパク質をコードする cDNA を pLexA ベクターの複数のクローニング部位に挿入し、連結して大腸菌に形質転換する。mRNA から調整した cDNA を、pB42AD ベクターの複数のクローニング部位に挿入して、結合させ、大腸菌に形質転換させて cDNA ライブラリを作製した。pLexA プラスミド及び pB42AD - cDNA ライブラリ作製物を大腸菌から単離し、これらを 2 : 1 の比率で用いて、polyethylene glycol / lithium acetate プロトコルに従ってコンピテント酵母 EGY48 [p8op-lacZ] 細胞を同時形質転換した。形質転換した酵母細胞をヒスチジン、トリプトファン、及びウラシルを含まない合成ドロップアウト (SD: synthetic dropout) 培地 (His, Trp, Ura) に移し、コロニーが成長してカウント出来るまで 30 でインキュベートした。このコロニーを 1 x TE (pH 7.5) の最小容量にプールし、2% アガロース (Gal)、1% ラフィノース (Raf)、及び 80 mg/ml の X-Gal (5-bromo-4-chloro-3-indoryl - b - D - galactoside) を加えた SD 培地 (His, Leu, Trp, Ura) に移し、青いコロニーの成長を調べた。発現したタンパク質と cDNA 融合タンパク質との間の相互作用により、EGY48 における LEU2 レポーター遺伝子の発現が活性化され、ロイシンを含まない培地 (Leu) でコロニーが成長した。また、相互作用により、p8op-lacZ レポーター作製物からの - ガラクトシダーゼの発現が活性化され、X-Gal 上で成長したコロニーにおいて青色コロニーが生成される。

【0148】

発現タンパク質と cDNA 融合タンパク質との間のポジティブな相互作用は、個々のポジティブコロニーを分離して、これらのコロニーを SD 液体培地 (Trp, Ura) で 30 で 2 日間成長させるて確認することができる。培養物のサンプルを SD 培地 (Trp, Ura) に移して、コロニーが現れるまで 30 でインキュベートする。このサンプルを、SD プレート (Trp, Ura) 及び SD プレート (His, Trp, Ura) 上でレプリカ培養する。ヒスチジン含有 SD 培地で成長するがヒスチジンを含まない培地では成長しないコロニーは、pLexA プラスミドが欠損している。ヒスチジンを必要とするコロニーを SD (Gal, Raf, X-Gal, Trp, Ura) 上で成長させ、白いコロニーを単離して成長させた。本哺乳動物タンパク質と物理的に相互作用するタンパク質をコードする cDNA を含む pB42AD - cDNA プラスミドを、酵母細胞から単離して特徴付けることが可能である。

【0149】

1.5 タンパク質活性の実証

DSCR1K1 またはその生物学的に活性な断片を、¹²⁵I - Bolton - Hunter 試薬 (Bolton A. E. (1973) Biochem. J. 133: 529 - 539) で標識する。マルチウェルプレートのウェルに予め配列しておいた候補リガンド分子を、標識した DSCR1K1 と共にインキュベートし、洗浄して、標識した DSCR1K1 複合体をアッセイする。様々な濃度の DSCR1K1 から得たデータを用いて、DSCR1K1 に結合した候補リガンド分子の数、親和性、会合についての値を計算する。

【0150】

本明細書に記載した全ての特許及び刊行物は、言及することを以って本明細書の一部とする。本発明の範囲及び概念から逸脱することなく本発明の方法及び装置の種々の改変が可能であることは当業者には明らかであろう。特定の好適な実施例に基づいて本発明を説明したが、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、分子生物学或いは関連する分野の専門家には、本明細書に記載の本発明の実施例の様々な改変は、特許請求の範囲に含まれることが容易に理解できよう。

【0151】

(表の簡単な説明)

表1は、SEQ ID NO: 1に相同なラット及びサルの核酸配列を示し、それらのヌクレオチド長、生物起源、SEQ ID NO: 1と重複する領域、及びSEQ ID NO: 1との同一性のパーセントを含む。

【図面の簡単な説明】

【図1A】

ヒトアミノ酸配列 (SEQ ID NO: 2) をコードするヒト核酸分子 (SEQ ID NO: 1) の領域を示す。このアラインメントは、MACDNASIS PROソフトウェア (Hitachi Software Engineering, South San Francisco CA) を用いて作成した。 10

【図1B】

ヒトアミノ酸配列 (SEQ ID NO: 2) をコードするヒト核酸分子 (SEQ ID NO: 1) の領域を示す。このアラインメントは、MACDNASIS PROソフトウェア (Hitachi Software Engineering, South San Francisco CA) を用いて作成した。

【図1C】

ヒトアミノ酸配列 (SEQ ID NO: 2) をコードするヒト核酸分子 (SEQ ID NO: 1) の領域を示す。このアラインメントは、MACDNASIS PROソフトウェア (Hitachi Software Engineering, South San Francisco CA) を用いて作成した。 20

【図2A】

ヒトDSCR1K1 (SEQ ID NO: 1)、ヒトDSCR1L1 (g1435040; SEQ ID NO: 10)、及びヒトDSCR1L2 (g6017919; SEQ ID NO: 11) の間の化学的及び構造的類似性を実証する。このアラインメントは、MEGALIGNプログラム (DNASTAR, Madison WI) を用いて作成した。

【図2B】

ヒトDSCR1K1 (SEQ ID NO: 1)、ヒトDSCR1L1 (g1435040; SEQ ID NO: 10)、及びヒトDSCR1L2 (g6017919; SEQ ID NO: 11) の間の化学的及び構造的類似性を実証する。このアラインメントは、MEGALIGNプログラム (DNASTAR, Madison WI) を用いて作成した。 30

【表1】

表1

核酸	インサイト配列 識別子	ヌクレオチド 長	生物名	SEQ ID No.1との オーバーラップ	SEQ ID No.1との 同一性 (%)
3	219435.1	1021	ラット	443-954 1040-1212	86 87
4	271176.1	193	ラット	1323-1461	91
5	291862.1	1348	ラット	2419-2517 2158-2299	94 86
6	299263.1	531	ラット	668-794 884-954	86 89
7	219652.1	233	ラット	1683-1765 1593-1644	89 90
8	206751.1	615	ラット	734-837	82
9	702110167H1	297	サル	218-274	94

10

20

30

40

【 1 A 】

```

190      199      208      217
5'  ATG TCA ACT TTG GAA AAA AAC AAC GTA GTG TNG
      M S T L R K N N V X
226      235      244      253      262      271
GGA ATG AGG GGA GAA TCA TAC TTC ATC GGA ATG AGG AGC CCA GGG CAG CAG GGA
G M R G E S Y F I G M R S P G Q Q G
280      289      307      316      325
CAC TTC CCA GAA GAT GGA GGA CTT TTC TTA CTG TCC TCC ATA GAC AGG GAC TGG
H V P E D G G L F L L C I D R D W
334      343      352      361      370      379
GCT GTC ACT CTT TTT GCA GAA GAA GCC TTT GAA GCA ATC ACT GAC TTC AAT
A V T R C F A E A F Q A I T D F N
388      387      406      415      424      433
GAC CTC CCC AAC TCG TTG TTT GCG TCC AAT GTT CAC CAG TCA GTC TTT GAA GGA
D L P N S L F A C N V H Q S V F E G
442      451      460      469      478      487
GAA GAG AGC AAG GAA AAA TTT GAG GGA CNG TTT CCG ACT TAT GAT GAC TGT GTG
E E S K E K F E G L F R T Y D D C V

```

FIGURE 1A

【 1 B 】

```

496      505      514      523      532      541
AGC TTC CAG CTA TTT TAG AGT TTC AGA CGT GTC CGT ATA AAC TTC AGC AAT CCG
T F Q L F K S F R R V R I N F S N P
550      559      568      577      586      595
AAA TCT GCA GCC CGA CTT AGG ATA GAG CTT CAT GAA ACC CAA TTC AGA GGG AAA
K S A A R A R I E L H E T Q F R G K
604      613      622      631      640      649
AAA TTA AAG CTC TAC TTT GCA CAG GTT CAG ACT CCA GAG ACA GAT GSA GAC AAA
K L K L Y F A Q V Q T P E T D G D K
658      667      676      685      694      703
CTG CAC TTG GCT CCA CCC CAG CTT GCC AAA CAG TTT CAC ATC TCC CCC CCG TCC
L H L A P P Q P A K Q F L I S P P S
712      721      730      739      748      757
TCC CCA CCT GTT GCG TGG CAG CCC ATC AAC GAT GCC ACG CCA GTC CTC AAC TAT
S F P V G W Q P I N D A T P V L N Y
766      775      784      793      802      811
GAC CTC CTC TAT GCT GTG GCC AAA CTA GGA CCA GGA GAG TAG TAT GAG CTC CTT
D L L Y A V A K L G P G E K Y E L H

```

FIGURE 1B

【 1 C 】

```

820      829      838      847      856      865
GCA GGG ACT GAG TCC ACC CCA AGT GTC GTC CAC GTG TCC GAC AGT GAC ATA
A G T E S T P S V V V H V C D S D I
874      883      892      901      910      919
GAG GAA GAA GAG GAC CCA AAG ACT TCC CCA AAG CCA AAA ATC ATC CAA ACT CCG
E E E D P K T S P K P K I I Q T R
928      937      946
CGT CCT GCC CTG CCA CCC TCC GTG TCC AAC 3'
R P G L P P S V S N

```

FIGURE 1C

【 2 A 】

```

1  M S T L F E R N M V V X G M R G E S Y F I G M R S P G Q Q H D S C R I L L G
1  M D C D V G I 435040
1  L R D T M K S W N D S Q S D 96017919
31 V P E D G G L F L L C C I D R D W A V T R C P A E E A F D S C R I L L G
6  L C S T D Q E E E E M I E G E N E D D L G I 435040
16  L C S T D Q E E E E M I E G E N E D D L 96017919
59 Q A I T D E N D L P N S L P A C N V H Q S V F E G E E S K E D S C R I L L G
37 D E M M D L S D L L P T S L F A C S V H E A V F E A R E I Q K E G I 435040
89 K F E G L F R T Y D D C V T F Q L F K S F R R V R I N F S N D S C R I L L G
67 R F E A L E T I Y D D Q V T F Q L F K S F R R V R I N F S K 96017919
119 P K S A A R A R I E L H E T Q F R G K K L K L Y F A Q V Q T D S C R I L L G
56 P K S A A R A R I E L H E T Q F R G K K L K L Y F A Q V Q T G I 435040
97 P E A A A R A R I E L H E T I E N I G I K L K L Y F A Q V Q M 96017919

```

FIGURE 2A

【 2 B 】

149 P E T D G D K L H L A P P O P A K Q F L I S P P S S P P V G D S C R L L I α
86 P E T D G D K L H L A P P O P A K Q F L I S P P S S P P V S G 91435040
127 S G E V R I D K S Y L I P P Q P V K Q F L I S P P A S P P V G 96017919
179 W O P I N D A T P V L N Y D L L Y A V A K L G P P G E K Y E L I D S C R L L I α
116 W P L I N D A T P V L N Y D L L Y A V A K L G P P G E K Y E L I 91435040
157 W K Q S E D A M P V I N Y D L L C A V I S K L G P P G E K Y E L I 96017919
209 H A G T E S T P S V V V H V C D S D I E E E E D P K T S P K D S C R L L I α
146 H A G T E S T P S V V V H V C D S D I E E E E D P K T S P K 91435040
187 H A G T E S T P S V V V H V C E S E T E E E E E T K - N P K 96017919
239 P K I I O T R R P G L P P S V S N
176 P K I I O T R R P G L P P S V S N
216 Q K I A Q T R R P D P E T A A L N E P Q T F D C A L

FIGURE 2B

【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
17 January 2002 (17.01.2002)

PCT

(10) International Publication Number
WO 02/04513 A2

- (51) International Patent Classification: C07K 14/47 (US). STREETER, David, G. [US/US]; 1211 Pinecrest Drive, Boulder Creek, CA 95006 (US).
- (21) International Application Number: PCT/US01/21982
- (22) International Filing Date: 11 July 2001 (11.07.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 09/614,474 11 July 2000 (11.07.2000) 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): LORING, Jeanne, F. [US/US]; 701 Santa Cruz Lane, Foster City, CA 94404 (US). TINGLEY, Debra, W. [US/US]; 119 Knollview Way, San Francisco, CA 94131 (US). EDWARDS, Carla, M. [US/US]; 41 Ventura Street, Half Moon Bay, CA 94019
- (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, 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, 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, BI, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, NI, NG, SN, TD, TG).

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

[Continued on next page]

(54) Title: DOWN SYNDROME CRITICAL REGION 1-LIKE 1 PROTEINS

```

          190       199       208       217
          ATG TCA ACT TTG GAA AAA AAC AAC GGA GTG TNG
          M S T L E K N N V V X
          -----
          226       235       244       253       262       271
GGA ATG AGG GGA GAA TCA TAC TTC ATC GGA ATG AGG AGC CCA GGG CAG CAG GGA
          G M R G E S Y F I G M H S P G Q Q G
          -----
          280       289       298       307       316       325
CAC GTC CCT GAA GAT GGA GGA CTT TTC TTA CTG TGC ATA GAC AGG GAC TGG
          H V P E D G G L F L L C C I D R D W
          -----
          334       343       352       361       370       379
GCT GTC ACT CGT TGT TTT GCA GAA GGC TTT CAA GCA ATC ACT GAC TTC AAT
          A V T R C P A E E A F Q A I T D F N
          -----
          388       397       406       415       424       433
GAC CTC CCC AAC TCG TTG TTT GCG TGC AAT GTT GAC CAG TCA GTG TTT GAA GGA
          D L P N S L F A C N V H Q S V P E G
          -----
          442       451       460       469       478       487
GAA GAG AGC AAG GAA AAA TTT GAG GGA CTG TTT CCG ACT TAT GAT GAC TCF GTG
          E E S K E K P E G L F R T Y D D C V

```

(57) Abstract: The invention provides a nucleic acid molecule and fragments thereof. It also provides for the use of the nucleic acid molecule for the characterization, diagnosis, evaluation, and treatment of conditions, diseases and disorders associated with gene expression and for the production of a model system. The invention additionally provides expression vectors and host cells for the production of the protein encoded by the nucleic acid molecule.

WO 02/04513 A2

WO 02/04513 A2



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

WO 02/04513

PCT/US01/21982

DOWN SYNDROME CRITICAL REGION 1-LIKE 1 PROTEINS**TECHNICAL FIELD**

This invention relates to a nucleic acid molecule which encodes a new mammalian protein
5 and to the use of these molecules in the characterization, diagnosis, and treatment of conditions such
as Alzheimer's disease, Down syndrome, and other forms of dementia.

BACKGROUND ART

Alzheimer's disease is a progressive neurodegenerative disorder that is characterized by the
formation of senile plaques and neurofibrillary tangles containing amyloid beta peptide. These
10 plaques are found in limbic and association cortices of the brain including hippocampus, temporal
cortices, cingulate cortex, amygdala, nucleus basalis, and locus caeruleus. Early in Alzheimer's
pathology, physiological changes are visible in the cingulate cortex (Minoshima *et al.* (1997) *Annals
of Neurology* 42:85-94). In subjects with advanced Alzheimer's disease, accumulating plaques
damage the neuronal architecture in limbic areas and eventually cripple the memory process.

15 Approximately twenty million people worldwide suffer with dementia that results from
Alzheimer's disease. The disease can be early onset (affecting individuals as young as 30 years of
age), familial, or sporadic. Familial Alzheimer's disease was once thought to be inherited strictly as
an autosomal dominant trait; however, this view is changing as more genetic determinants are
isolated. For example, some normal allelic variants of apolipoprotein E (ApoE), which is found in
20 senile plaques, can either protect against or increase the risk of developing the disease (Strittmatter *et
al.* (1993) *Proc Natl Acad Sci* 90:1977-1981).

Mutations in four genes are known to predispose an individual to Alzheimer's disease:
ApoE, amyloid precursor protein (APP), presenilin-1, and presenilin-2 (Selkoe (1999) *Nature*
399:A23-A31). The e4 allele of the ApoE gene confers increased risk for late onset Alzheimer's
25 disease. β -amyloid protein (A β) is the major component of senile plaques, and it is normally formed
when β - and γ - secretases cleave APP. In Alzheimer's disease patients, large quantities of A β are
generated and accumulate extracellularly in these neuropathological plaques.

Associations between Alzheimer's disease and many other genes and proteins have been
reported. Fetal Alzheimer antigen and synuclein α are found in brain plaques and tangles.
30 Inheritance of some gene polymorphisms is also linked to increased risk of developing the disease.
For example, a polymorphism in the gene encoding β 2-macroglobulin, a protein that can act as a
protease inhibitor, is associated with increased risk for developing a late-onset form of Alzheimer's
disease.

Experiments using microarray technology have provided additional evidence for changes in
35 expression of specific genes in tissues from subjects with Alzheimer's disease. One such gene, Down

WO 02/04513

PCT/US01/21982

syndrome critical region 1-like 1 (Mazowiecki *et al.* (1996) *J Biol Chem* 271:14567-14571) also named DSCR1L1 (g1435040), was found to be down-regulated more than two-fold in Alzheimer's tissue. DSCR1L1 encodes a thyroid hormone responsive protein and is a member of a gene family that includes DSCR1 (g7657042) and DSCR1L2 (g6017918). The first member of the family,

5 DSCR1, was cloned and named based on its proximity to the Down syndrome region of chromosome 21, but the function of the encoded protein and its role in Down syndrome, if any, remains unclear. The defining motifs of this family of proteins include an N-terminal RNA-binding domain, which is similar to those found in many RNA-binding proteins and in some single-stranded DNA-binding proteins, and a central short, unique serine-proline motif that includes an ISPPXSPP box that may be
10 a target for phosphorylation (Fuentes *et al.* (1995) *Hum Mol Genet* 4:1935-1944).

Based on general features of the amino acid sequences, the DSCR1 genes are likely to encode proteins involved in transcriptional regulation and signal transduction (Strippoli *et al.* (2000) *Genomics* 64:252-263). The three genes appear to be differentially expressed: DSCR1 is highly expressed in fetal brain and heart, DSCR1L1 is evident in heart, brain, liver and skeletal muscle, and
15 DSCR1L2 is expressed in most tissues including blood.

There are several connections between Alzheimer's disease and Down syndrome, such as the appearance of enhanced β -amyloid deposits in middle-aged Down syndrome patients. The thyroid hormone responsive properties of DSCR1L1 may be relevant to the role for this gene in Alzheimer's disease and/or Down syndrome: both Alzheimer's and Down's patients have been shown to have
20 disturbances in thyroid hormone metabolism including elevated antithyroglobulin antibodies and reduced levels of prealbumin in the cerebrospinal fluid of Alzheimer's patients and pronounced subclinical hypothyroidism in Down's patients who have clinical manifestations of Alzheimer's disease (Sutherland *et al.* (1992) *Neurobiol Aging* 13:301-312). Conversely, hypothyroid patients display some of the same neurologic symptoms seen in Alzheimer's patients, and a history of thyroid
25 dysfunction is considered to be a risk factor for developing Alzheimer's disease. One possible connection is the discovery that thyroid hormone negatively regulates the transcriptional activity of the APP gene (Belandia *et al.* (1998) *J Biol Chem* 273:30366-30371); over-expression of APP has been hypothesized to cause neuronal degeneration by a mechanism involving increased production of β -amyloid protein. The down-regulation of the thyroid responsive DSCR1L1 gene, and possibly
30 other DSCR1 genes, in Alzheimer's patients also provides a link between Alzheimer's and thyroid hormone function.

The discovery of additional nucleic acid molecules encoding members of the DSCR1 protein family may be used for the diagnosis, prognosis or treatment of Alzheimer's disease, Down syndrome and other forms of dementia.

35

WO 02/04513

PCT/US01/21982

SUMMARY OF THE INVENTION

The invention is based on the discovery of a purified nucleic acid molecule encoding DSCRIL1 α protein which satisfies a need in the art by providing compositions useful in the characterization, diagnosis, and treatment of conditions such as Alzheimer's disease, Down syndrome
5 and other forms of dementia.

The invention provides an isolated nucleic acid molecule or a fragment thereof encoding a protein or a portion thereof selected from the group consisting of an amino acid sequence of SEQ ID NO:2, a variant having at least 95% identity to the amino acid sequence of SEQ ID NO:2, and an antigenic epitope of SEQ ID NO:2. The invention also provides an isolated mammalian nucleic acid
10 molecule or the complement thereof selected from the group consisting of a nucleic acid sequence of SEQ ID NO:1, a variant cDNA having at least 85% identity to the nucleic acid sequence of SEQ ID NO:1, and a fragment of SEQ ID NO:1. The invention additionally provides a composition, a substrate, and a probe or target comprising the nucleic acid molecule, or the complement of the nucleic acid molecule, encoding DSCRIL1 α . The invention further provides a vector containing the
15 nucleic acid molecule, a host cell containing the vector and a method for using the nucleic acid molecule to make DSCRIL1 α . The invention still further provides a transgenic cell line or organism comprising the vector containing the nucleic acid molecule encoding DSCRIL1 α . The invention additionally provides a mammalian fragment or the complement thereof selected from the group consisting of SEQ ID NOs:3-9. In one aspect, the invention provides a substrate containing a nucleic
20 acid molecule. In a second aspect, the substrate is used in methods of detection, screening, and purification. In a further aspect, nucleic acid molecule is a single stranded complementary RNA or DNA molecule.

The invention provides a method for using a nucleic acid molecule to detect the differential expression of a nucleic acid in a sample comprising hybridizing the nucleic acid molecule to the
25 nucleic acids, thereby forming hybridization complexes and comparing hybridization complex formation with a standard, wherein the comparison indicates the differential expression of the nucleic acids in the sample. In one aspect, the method of detection further comprises amplifying the nucleic acids of the sample prior to hybridization. In another aspect, differential expression of the nucleic acid molecule when compared to a standard is used to diagnose Alzheimer's disease, Down syndrome
30 and other forms of dementia. In another aspect, the nucleic acid molecule or a fragment or a complement thereof may comprise an element on an array.

The invention additionally provides a method for using a nucleic acid molecule or a fragment or a complement thereof to screen a library or plurality of molecules or compounds to identify at least
35 one ligand which specifically binds the nucleic acid molecule, the method comprising combining the nucleic acid molecule with the molecules or compounds under conditions allowing specific binding.

WO 02/04513

PCT/US01/21982

and detecting specific binding to the nucleic acid molecule, thereby identifying a ligand which specifically binds the nucleic acid molecule. In one aspect, the molecules or compounds are selected from DNA molecules, RNA molecules, peptide nucleic acids, artificial chromosome constructions, peptides, transcription factors, repressors, and regulatory molecules.

5 The invention provides a purified protein or a portion thereof selected from the group consisting of an amino acid sequence of SEQ ID NO:2, a variant having 95% identity to the amino acid sequence of SEQ ID NO:2, an antigenic epitope of SEQ ID NO:2, an oligopeptide of SEQ ID NO:2, and a biologically active portion of SEQ ID NO:2. The invention also provides a composition comprising the purified protein and a pharmaceutical carrier. The invention still further provides a
10 method for using a protein to screen a library or a plurality of molecules or compounds to identify at least one ligand, the method comprising combining the protein with the molecules or compounds under conditions to allow specific binding and detecting specific binding, thereby identifying a ligand which specifically binds the protein. In one aspect, the molecules or compounds are selected from DNA molecules, RNA molecules, peptide nucleic acids, peptides, proteins, mimetics, agonists,
15 antagonists, antibodies, immunoglobulins, inhibitors, and drugs. In another aspect, the ligand is used to treat a subject with Alzheimer's disease, Down syndrome or other forms of dementia.

The invention provides a method of using a mammalian protein to screen a subject sample for antibodies which specifically bind the protein comprising isolating antibodies from the subject sample, contacting the isolated antibodies with the protein under conditions that allow specific
20 binding, dissociating the antibody from the bound-protein, and comparing the quantity of antibody with known standards, wherein the presence or quantity of antibody is diagnostic of Alzheimer's disease, Down syndrome and other forms of dementia. The invention also provides a method of using a mammalian protein to prepare and purify antibodies comprising immunizing an animal with the protein under conditions to elicit an antibody response, isolating animal antibodies, attaching the
25 protein to a substrate, contacting the substrate with isolated antibodies under conditions to allow specific binding to the protein, dissociating the antibodies from the protein, thereby obtaining purified antibodies.

The invention provides a purified antibody which bind specifically to DSCR1L1 α . The invention also provides a method of using an antibody to diagnose Alzheimer's disease, Down
30 syndrome and other forms of dementia comprising combining the antibody with a sample under conditions to allow binding, comparing the amount of bound antibody with known standards, thereby diagnosing the presence or absence of the disease. The invention further provides a method of using an antibody to treat Alzheimer's disease, Down syndrome and other forms of dementia comprising administering to a patient in need of such treatment a pharmaceutical composition comprising the
35 purified antibody.

WO 02/04513

PCT/US01/21982

The invention provides a method for inserting a marker gene into the genomic DNA of a mammal to disrupt the expression of the endogenous polynucleotide. The invention also provides a method for using a nucleic acid molecule to produce a mammalian model system, the method comprising constructing a vector containing the nucleic acid molecule selected from SEQ ID NOs:1 and 3-9, transforming the vector into an embryonic stem cell, selecting a transformed embryonic stem, microinjecting the transformed embryonic stem cell into a mammalian blastocyst, thereby forming a chimeric blastocyst, transferring the chimeric blastocyst into a pseudopregnant dam, wherein the dam gives birth to a chimeric offspring containing the nucleic acid molecule in its germ line, and breeding the chimeric mammal to produce a homozygous, mammalian model system.

BRIEF DESCRIPTION OF THE FIGURES AND TABLE

Figures 1A, 1B, and 1C show the region of the human nucleic acid molecule (SEQ ID NO:1) encoding the human amino acid sequence (SEQ ID NO:2). The alignment was produced using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA).

Figures 2A and 2B demonstrate the chemical and structural similarity among the human DSCR1L1 α (SEQ ID NO:1), human DSCR1L1 (g1435040;SEQ ID NO:10), and human DCSR1L2 (g6017919;SEQ ID NO:11). The alignment was produced using the MEGALIGN program (DNASTAR, Madison WI).

Table 1 shows rat and monkey nucleic acid sequences which have homology with SEQ ID NO:1 and includes their nucleotide length, biological source, region of overlap with SEQ ID NO:1, and percent identity with SEQ ID NO:1.

DESCRIPTION OF THE INVENTION

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

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

35

WO 02/04513

PCT/US01/21982

Definitions

"Down syndrome critical region 1-like 1 α (DSCR1L1 α) protein" refers to a purified protein obtained from any mammalian species, including bovine, canine, murine, opine, porcine, rodent, simian, and preferably the human species, and from any source, whether natural, synthetic, 5 semi-synthetic, or recombinant.

"Array" refers to an ordered arrangement of at least two nucleic acid molecules on a substrate. At least one of the nucleic acid molecules represents a control or standard sequence, and the other, a nucleic acid molecule of diagnostic interest. The arrangement of from about two to about 40,000 nucleic acid molecules on the substrate assures that the size and signal intensity of each 10 labeled hybridization complex formed between a nucleic acid molecule and a sample nucleic acid is individually distinguishable.

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

15 "Nucleic acid molecule" refers to an isolated polynucleotide, cDNA, or any fragment or complement thereof. It may have originated recombinantly or synthetically, be double-stranded or single-stranded, represent coding and/or noncoding sequence, an exon with or without an intron from a genomic DNA molecule.

The phrase "nucleic acid molecule encoding a protein" refers to a nucleic acid sequence that 20 closely aligns with sequences which encode conserved regions, motifs or domains that were identified by employing analyses well known in the art. These analyses include BLAST (Basic Local Alignment Search Tool; Altschul (1993) *J Mol Ecol* 36: 290-300; Altschul *et al.* (1990) *J Mol Biol* 215:403-410) which provides identity within the conserved region. Brenner *et al.* (1998; *Proc Natl Acad Sci* 95:6073-6078) who analyzed BLAST for its ability to identify structural homology by 25 sequence identity found 30% identity is a reliable threshold for sequence alignments of at least 150 residues and 40% is a reasonable threshold for alignments of at least 70 residues (Brenner *et al.*, page 6076, column 2).

"Derivative" refers to a nucleic acid molecule or a protein that has been subjected to a chemical modification. Derivatization of a nucleic acid molecule can involve substitution of a 30 nontraditional base such as queosine or of an analog such as hypoxanthine. These substitutions are well known in the art. Derivatization of a protein involves the replacement of a hydrogen by an acetyl, acyl, alkyl, amino, formyl, or morpholino group. Derivative molecules retain the biological activities of the naturally occurring molecules but may confer advantages such as longer lifespan or enhanced activity.

35 "Differential expression" refers to an increased or up-regulated or a decreased or down-

WO 02/04513

PCT/US01/21982

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.

"Disorder" refers to conditions, diseases or syndromes in which the nucleic acid molecules and DSCR1 proteins are differentially expressed.

5 "Fragment" refers to a chain of consecutive nucleotides from about 200 to about 700 base pairs in length. Fragments may be used in PCR or hybridization technologies to identify related nucleic acid molecules and in binding assays to screen for a ligand. Nucleic acids and ligands identified in this manner are useful as therapeutics to regulate replication, transcription or translation.

A "hybridization complex" is formed between a nucleic acid molecule and a nucleic acid of a 10 sample when the purine of one molecule hydrogen bond with the pyrimidine of the complementary molecule, e.g., 5'-A-G-T-C-3' base pairs with 3'-T-C-A-G-5'. The degree of complementarity and the use of nucleotide analogs affect the efficiency and stringency of hybridization reactions.

"Ligand" refers to any agent, molecule, or compound which will bind specifically to a 15 complementary site on a polynucleotide or an epitope of a protein. Such ligands stabilize or modulate the activity of polynucleotides or proteins and may be composed of inorganic or organic substances including nucleic acids, proteins, carbohydrates, fats, and lipids.

"Oligonucleotide" refers a single stranded molecule from about 18 to about 60 nucleotides in 20 length which may be used in hybridization or amplification technologies or in regulation of replication, transcription or translation. Equivalent terms include amplimer, primer, and oligomer.

"Portion" refers to any part of a protein used for any purpose; but especially, to an epitope for 25 the screening of ligand or for the production of antibodies.

"Post-translational modification" of a protein can involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and the like. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cellular location, cell 25 type, pH, enzymatic milieu, and the like.

"Probe" refers to a nucleic acid molecule that hybridizes to at least one nucleic acid in a 30 sample. Where targets are single stranded, probes are complementary single strands. Probes can be labeled with reporter molecules for use in hybridization reactions including Southern, northern, in situ, dot blot, array, and like technologies or in screening assays.

"Protein" refers to a polypeptide or any portion thereof. A "portion" of a protein refers to that 35 length of amino acid sequence which would retain at least one biological activity, a domain identified by PFAM or PRINTS analysis or an antigenic epitope of the protein identified using Kate-Doolittle algorithms of the PROTEAN program (DNASTAR, Madison WI). An "oligopeptide" is an amino acid sequence from about five residues to about 15 residues that is used as part of a fusion protein to produce an antibody.

WO 02/04513

PCT/US01/21982

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

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

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

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

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

"Variant" refers to molecules that are recognized variations of a nucleic acid molecule or a protein encoded by the nucleic acid molecule. 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 nucleic acid molecule and may differ by about three bases per hundred bases.

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

THE INVENTION

The invention is based on the discovery of new mammalian nucleic acid molecules which encode the mammalian protein DSCR1L1 α , and on the use of the nucleic acid molecule, or fragments thereof, and protein, or portions thereof, as compositions in the characterization, diagnosis, and treatment of conditions such as Alzheimer's disease, Down syndrome and other forms of dementia.

WO 02/04513

PCT/US01/21982

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:2 as shown in Figures 1A, 1B, and 1C. DSCRIL1 α has a single-stranded nucleic acid-binding domain characterized by positively charged and aromatic amino acids from residue 129 to residue 199, which is similar to known RNA recognition motifs. In addition,

5 DSCRIL1 α contains the ISPPXSPP box characteristic of DSCR1 proteins from residue 169 to residue 176. As shown in Figures 2A and 2B, DSCRIL1 α shares chemical and structural homology with human DSCRIL1 (g1435040; SEQ ID NO:10), and human DCSRIL2 (g6017919; SEQ ID NO:11). In particular, DSCRIL1 α shares 90.3% identity with DSCRIL1 and 57.3% identity with DSCRIL2. All three proteins share the single-stranded nucleic acid-binding domain and the

10 ISPPXSPP box. SEQ ID NO:1 differs from the gene encoding DSCRIL1 (g1435039; SEQ ID NO:12) and the gene encoding DSCRIL2 (g6017918; SEQ ID NO:13) in the 5' untranslated region, and is further distinguished from these and other genes encoding DSCR1 family proteins by the first 70 amino acids encoded by the open reading frame from about amino acid residue M1 to about amino acid residue S70 of SEQ ID NO:1. A probe from the region of SEQ ID NO:1 from about nucleotide

15 1 to about nucleotide 400 would be useful for identifying naturally occurring molecules encoding DSCRIL1 α , allelic variants, or related molecules. An antigenic epitope from the region of DSCRIL1 α from about amino acid residue M1 to about amino acid residue S70 would be useful for the production of antibodies to DSCRIL1 α which would distinguish between DSCRIL1 α and similar proteins.

20 Electronic northern analysis shows the expression of this sequence in various libraries, with the highest abundance in tissues from the nervous system, including tissues associated with schizophrenia, Huntington's disease, epilepsy, and amyotrophic lateral sclerosis. Consistent with its observed 2-fold down-regulation on microarrays hybridized with samples from subjects with diagnosed Alzheimer's disease (USSN 09/534,846, filed May 5, 2000, hereby incorporated by

25 reference), is the absence of DSCRIL1 α expression in 7 of 8 libraries from subjects with Alzheimer's disease.

Table 1 shows nucleic acid fragments from rat and monkey and their sequence coverage and identity with SEQ ID NO:1. Columns 1 and 2 list the SEQ ID NO and Incyte sequence identifier, respectively, for each nucleic acid sequence. Column 3 lists the nucleotide length for each fragment,

30 Column 4, the range of nucleotide residues in SEQ ID NO:1 over which each fragment shows identity with SEQ ID NO:1, Column 5 identifies the source of the organism, and Column 6 shows percent identity between each fragment and SEQ ID NO:1 over the nucleotide range set forth in column 4.

These nucleic acid molecules are particularly useful for producing transgenic cell lines or

WO 02/04513

PCT/US01/21982

organisms which model human disorders and upon which potential therapeutic treatments for such disorders may be tested. Of particular note is the expression of SEQ ID NO:9 in cingulate cortex, an area that shows functional deficits early in the Alzheimer's disease process. This monkey nucleic acid molecule would be useful in a model system of Alzheimer's disease. The monkey nucleic acid molecule may also be used as a probe to search for homologous human DSCR1L1 gene family sequences that similarly have expression restricted to cingulate cortex.

Characterization and Use of the Invention

cDNA libraries

In a particular embodiment disclosed herein, mRNA was isolated from mammalian cells and tissues using methods which are well known to those skilled in the art and used to prepare the cDNA libraries. The Incyte clones listed above were isolated from mammalian cDNA libraries. At least one library preparation representative of the invention is described in the EXAMPLES below. The consensus mammalian sequences were chemically and/or electronically assembled from fragments including Incyte clones and extension and/or shotgun sequences using computer programs such as Phrap (P. Green, University of Washington, Seattle WA), GELVIEW Fragment Assembly system (Genetics Computer Group, Madison WI), and AUTOASSEMBLER application (Applied Biosystems, Foster City CA).

Sequencing

Methods for sequencing nucleic acids are well known in the art and may be used to practice any of the embodiments of the invention. These methods employ enzymes such as the Klenow fragment of DNA polymerase I, SEQUENASE, Taq DNA polymerase and thermostable T7 DNA polymerase (Amersham Pharmacia Biotech (APB), Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Rockville MD). Preferably, sequence preparation is automated with machines such as the HYDRA microdispenser (Robbins Scientific, Sunnyvale CA), MICROLAB 2200 (Hamilton, Reno NV), and the DNA ENGINE thermal cycler (PTC200; MJ Research, Watertown MA). Machines used for sequencing include the ABI 3700, 377 or 373 DNA sequencing systems (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (APB), and the like. The sequences may be analyzed using a variety of algorithms which are well known in the art and described in Ausubel (1997; Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7) and Meyers (1995; Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

Shotgun sequencing is used to generate more sequence from cloned inserts derived from multiple sources. Shotgun sequencing methods are well known in the art and use thermostable DNA polymerases, heat-labile DNA polymerases, and primers chosen from representative regions flanking

WO 02/04513

PCT/US01/21982

the nucleic acid molecules of interest. Prefinished sequences (incomplete assembled sequences) are inspected for identity using various algorithms or programs such as CONSED (Gordon (1998) Genome Res. 8:195-202) which are well known in the art. Contaminating sequences including vector or chimeric sequences or deleted sequences can be removed or restored, respectively, organizing the prefinished sequences into finished sequences.

Extension of a Nucleic Acid Sequence

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

USE OF THE MAMMALIAN NUCLEIC ACID MOLECULE

Hybridization

The mammalian nucleic acid molecule and fragments thereof can be used in hybridization technologies for various purposes. A probe may be designed or derived from unique regions such as the 5' regulatory region or from a conserved motif such as the N-terminal RNA-binding domain found in the DSCR1 protein family and used in protocols to identify naturally occurring molecules encoding the mammalian protein, allelic variants, or related molecules. The probe may be DNA or RNA, is usually single stranded and should have at least 50% sequence identity to any of the nucleic acid sequences. Hybridization probes may be produced using oligolabeling, nick translation, end-labeling, or PCR amplification in the presence of labeled nucleotide. A vector containing the nucleic acid molecule or a fragment thereof may be used to produce an mRNA probe *in vitro* by addition of an RNA polymerase and labeled nucleotide. These procedures may be conducted using commercially available kits such as those provided by APB.

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

WO 02/04513

PCT/US01/21982

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

10 Microarrays may be prepared and analyzed using methods known in the art. Oligonucleotides may be used as either probes or targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and single nucleotide polymorphism. Such information may be used to determine gene function; to understand the genetic basis of a condition, disease, or disorder; to

15 diagnose a condition, disease, or disorder; and to develop and monitor the activities of therapeutic agents. (See, e.g., Brenman *et al.* (1995) USPN 5,474,796; Schena *et al.* (1996) Proc Natl Acad Sci 93:10614-10619; Baldeschweiler *et al.* (1995) PCT application WO95/251116; Shalon *et al.* (1995) PCT application WO95/35505; Heller *et al.* (1997) Proc Natl Acad Sci 94:2150-2155; and Heller *et al.* (1997) USPN 5,605,662.)

20 Hybridization probes are also useful in mapping the naturally occurring genomic sequence. The probes may be hybridized to: 1) a particular chromosome; 2) a specific region of a chromosome; 3) artificial chromosome constructions such as human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), or bacterial artificial chromosomes (BACs), bacterial P1 constructions; or 5) cDNA libraries made from any of these sources.

25 Expression

A multitude of nucleic acid molecules encoding the mammalian DSCR1L1 α protein may be cloned into a vector and used to express the protein, or portions thereof, in host cells. The nucleic acid sequences can be engineered by such methods as DNA shuffling (Stammer and Cramer (1996) USPN 5,830,721) and site-directed mutagenesis to create new restriction sites, alter glycosylation

30 patterns, change codon preference to increase expression in a particular host, produce splice variants, extend half-life, and the like. The expression vector may contain transcriptional and translational control elements (promoters, enhancers, specific initiation signals, and polyadenylated 3' sequence) from various sources which have been selected for their efficiency in a particular host. The vector, nucleic acid molecule, and regulatory elements are combined using *in vitro* recombinant DNA

35 techniques, synthetic techniques, and/or *in vivo* genetic recombination techniques well known in the

WO 02/04513

PCT/US01/21982

art and described in Sambrook (supra, ch. 4, 8, 16 and 17).

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

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

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

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

35 Recovery of Proteins from Cell Culture

WO 02/04513

PCT/US01/21982

Heterologous moieties engineered into a vector for ease of purification include glutathione S-transferase (GST), calmodulin binding peptide (CBP), 6-His, FLAG, MYC, and the like. GST, CBP, and 6-His are purified using commercially available affinity matrices such as immobilized glutathione, calmodulin, and metal-chelate resins, respectively. FLAG and MYC are purified using commercially available monoclonal and polyclonal antibodies. A proteolytic cleavage site may be located between the desired protein sequence and the heterologous moiety for ease of separation following purification. Methods for recombinant protein expression and purification are discussed in Ausubel (supra, unit 16) and are commercially available.

Chemical Synthesis of Peptides

Proteins or portions thereof may be produced not only by recombinant methods, but also by using chemical methods well known in the art. Solid phase peptide synthesis may be carried out in a batchwise or continuous flow process which sequentially adds α -amino- and side chain-protected amino acid residues to an insoluble polymeric support via a linker group. A linker group such as methylamine-derivatized polyethylene glycol is attached to poly(styrene-co-divinyl benzene) to form the support resin. The amino acid residues are *N*- α -protected by acid labile Boc (t-butyloxycarbonyl) or base-labile Fmoc (9-fluorenylmethoxycarbonyl). The carboxyl group of the protected amino acid is coupled to the amine of the linker group to anchor the residue to the solid phase support resin. Trifluoroacetic acid or piperidine are used to remove the protecting group in the case of Boc or Fmoc, respectively. Each additional amino acid is added to the anchored residue using a coupling agent or pre-activated amino acid derivative, and the resin is washed. The full length peptide is synthesized by sequential deprotection, coupling of derivitized amino acids, and washing with dichloromethane and/or *N,N*-dimethylformamide. The peptide is cleaved between the peptide carboxyl terminus and the linker group to yield a peptide acid or amide. (Novabiochem 1997/98 Catalog and Peptide Synthesis Handbook, San Diego CA pp. S1-S20). Automated synthesis may also be carried out on machines such as the ABI 431A peptide synthesizer (Applied Biosystems). A protein or portion thereof may be purified by preparative high performance liquid chromatography and its composition confirmed by amino acid analysis or by sequencing (Creighton (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY).

Preparation and Screening of Antibodies

Various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with mammalian DSCRIL1 α protein or any portion thereof. Adjuvants such as Freund's, mineral gels, and surface active substances such as lysolecithin, plutonic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemacyanin (KLH), and dinitrophenol may be used to increase immunological response. The oligopeptide, peptide, or portion of protein used to induce antibodies should consist of at least about five amino acids, more preferably ten amino acids, which

WO 02/04513

PCT/US01/21982

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

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

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

The mammalian DSCRIL1 α protein or a portion thereof may be used in screening assays of phagemid or B-lymphocyte immunoglobulin libraries to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoassays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between the protein and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes is preferred, but a competitive binding assay may also be employed (Pound (1998) *Immunochemical Protocols*, Humana Press, Totowa NJ).

25 Labeling of Molecules for Assay

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid, amino acid, and antibody assays. Synthesis of labeled molecules may be achieved using Promega (Madison WI) or APB kits for incorporation of a labeled nucleotide such as ³²P-dCTP, Cy3-dCTP or Cy5-dCTP or amino acid such as ³⁵S-methionine (APB).

30 Nucleotides and amino acids may be directly labeled with a variety of substances including fluorescent, chemiluminescent, or chromogenic agents, and the like, by chemical conjugation to amines, thiols and other groups present in the molecules using reagents such as BIODIPY or FITC (Molecular Probes, Eugene OR).

DIAGNOSTICS

35 The nucleic acid molecules, fragments, oligonucleotides, complementary RNA and DNA

WO 02/04513

PCT/US01/21982

molecules, and PNAs may be used to detect and quantify altered gene expression, absence/presence vs. excess, expression of mRNAs or to monitor mRNA levels during therapeutic intervention. Conditions, diseases or disorders associated with differential expression include: Alzheimer's disease, dementia, Down syndrome, Huntington's disease and Pick's disease. The diagnostic assay may use
5 hybridization or amplification technology to compare gene expression in a biological sample from a patient to standard samples in order to detect altered gene expression. Qualitative or quantitative methods for this comparison are well known in the art.

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

In order to provide a basis for the diagnosis of a condition, disease or disorder associated
15 with gene expression, a normal or standard expression profile is established. This may be accomplished by combining a biological sample taken from normal subjects, either animal or human, with a probe under conditions for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained using normal subjects with values from an experiment in which a known amount of a purified target sequence is used. Standard values obtained in this manner
20 may be compared with values obtained from samples from patients who are symptomatic for a particular condition, disease, or disorder. Deviation from standard values toward those associated with a particular condition is used to diagnose that condition.

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

Immunological Methods

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

WO 02/04513

PCT/US01/21982

Pound, *supra*.)**THERAPEUTICS**

Chemical and structural similarity, e.g., in the context of the RNA-binding domain and the conserved serine/proline motif, exist between regions of the SEQ ID NO:2 and other DSCR1 family proteins such as human DSCR1, human DSCR1L1 (g1435040), and human DSCR1L2 (g6017919). In addition, gene expression is closely associated with nervous system tissues and appears to play a role in Alzheimer's disease and Down syndrome. In the treatment of conditions associated with increased expression or activity, it is desirable to decrease expression or protein activity. In the treatment of conditions associated with decreased expression or activity, it is desirable to increase expression or protein activity.

In one embodiment, the mammalian protein or a portion or derivative thereof may be administered to a subject to treat or prevent a condition associated with altered expression or activity of the mammalian protein. Examples of such conditions include, but are not limited to, Alzheimer's disease, dementia, Down syndrome, Huntington's disease and Pick's disease.

In another embodiment, a pharmaceutical composition comprising a purified mammalian protein in conjunction with a pharmaceutical carrier may be administered to a subject to treat or prevent a condition associated with altered expression or activity of the endogenous protein including, but not limited to, those provided above.

In a further embodiment, a ligand which modulates the activity of the mammalian protein may be administered to a subject to treat or prevent a condition associated with altered lifespan, expression, or activity of the protein including, but not limited to, those listed above. In one aspect, an antibody which specifically binds the mammalian protein may be used as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express the mammalian protein.

In an additional embodiment, a vector capable of expressing the mammalian protein or a portion or derivative thereof may be administered to a subject to treat or prevent a condition associated with altered lifespan, expression, or activity of protein including, but not limited to, those described above.

In a still further embodiment, a vector expressing the complement of the nucleic acid molecule or fragments thereof may be administered to a subject to treat or prevent a condition associated with altered lifespan, expression, or activity of the protein including, but not limited to, those described above.

Any of the nucleic acid molecules, complementary molecules and fragments thereof, proteins or portions thereof, vectors delivering these nucleic acid molecules or proteins, and their ligand may be administered in combination with other therapeutic agents. Selection of the agents for use in

WO 02/04513

PCT/US01/21982

combination therapy may be made by one of ordinary skill in the art according to conventional pharmaceutical principles. A combination of therapeutic agents may act synergistically to effect treatment of a particular condition at a lower dosage of each agent.

Modification of Gene Expression Using Nucleic Acids

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

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

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

30 Screening Assays

The nucleic acid molecule encoding the mammalian protein may be used to screen a library of molecules for specific binding affinity. The libraries may be DNA molecules, RNA molecules, PNAs, peptides, proteins such as transcription factors, enhancers, repressions, and other ligand which regulate the activity, replication, transcription, or translation of the nucleic acid molecule in the
35 biological system. The assay involves combining the mammalian nucleic acid molecule or a

WO 02/04513

PCT/US01/21982

fragment thereof with the library of molecules under conditions allowing specific binding, and detecting specific binding to identify at least one molecule which specifically binds the nucleic acid molecule.

Similarly the mammalian protein or a portion thereof may be used to screen libraries of 5 molecules in any of a variety of screening assays. The portion of the protein employed in such screening may be free in solution, affixed to an abiotic or biotic substrate (e.g. borne on a cell surface), or located intracellularly. Specific binding between the protein and molecule may be measured. Depending on the kind of library being screened, the assay may be used to identify DNA, RNA, or PNA molecules, agonists, antagonists, antibodies, immunoglobulins, inhibitors, peptides, 10 proteins, drugs, or any other ligand, which specifically binds the protein. One method for high throughput screening using very small assay volumes and very small amounts of test compound is described in USPN 5,876,946, which screens large numbers of molecules for enzyme inhibition or receptor binding.

Purification of Ligand

15 The nucleic acid molecule or a fragment thereof may be used to purify a ligand from a sample. A method for using a mammalian nucleic acid molecule or a fragment thereof to purify a ligand would involve combining the nucleic acid molecule or a fragment thereof with a sample under conditions to allow specific binding, detecting specific binding, recovering the bound protein, and using an appropriate agent to separate the nucleic acid molecule from the purified ligand.

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

Pharmacology

25 Pharmaceutical compositions are those substances wherein the active ingredients are contained in an effective amount to achieve a desired and intended purpose. The determination of an effective dose is well within the capability of those skilled in the art. For any compound, the therapeutically effective dose may be estimated initially either in cell culture assays or in animal 30 models. The animal model is also used to achieve a desirable concentration range and route of administration. Such information may then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of protein or inhibitor which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity of such agents may be 35 determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g.,

WO 02/04513

PCT/US01/21982

ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index, and it may be expressed as the ratio, LD50/ED50. Pharmaceutical compositions which exhibit large therapeutic indexes are preferred. The data obtained from cell culture assays and animal studies are used in formulating a range of dosage for human use.

MODEL SYSTEMS

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

15 Toxicology

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

Genetic toxicology identifies and analyzes the ability of an agent to produce genetic mutations. Genotoxic agents usually have common chemical or physical properties that facilitate interaction with nucleic acids and are most harmful when chromosomal aberrations are passed along to progeny. Toxicological studies may identify agents that increase the frequency of structural or functional abnormalities in progeny if administered to either parent before conception, to the mother during pregnancy, or to the developing organism. Mice and rats are most frequently used in these tests because of their short reproductive cycle which produces the number of organisms needed to satisfy statistical requirements.

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

Prolonged toxicity tests are based on the repeated administration of the agent. Rat and dog are commonly used in these studies to provide data from species in different families. With the exception of carcinogenesis, there is considerable evidence that daily administration of an agent at

WO 02/04513

PCT/US01/21982

high-dose concentrations for periods of three to four months will reveal most forms of toxicity in adult animals.

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

Transgenic Animal Models

Transgenic rodents which over-express or under-express a gene of interest may be inbred and used to model human diseases or to test therapeutic or toxic agents. (See USPN 4,736,866; USPN 5,175,383; and USPN 5,767,337). In some cases, the introduced gene may be activated at a specific time in a specific tissue type during fetal development or postnatal. Expression of the transgene is monitored by analysis of phenotype or tissue-specific mRNA expression, in transgenic animals before, during, and after being challenged with experimental drug therapies.

Embryonic Stem Cells

Embryonic stem cells (ES) isolated from rodent embryos retain the potential to form an embryo. When ES cells are placed inside a carrier embryo, they resume normal development and contribute to all tissues of the live-born animal. ES cells are the preferred cells used in the creation of experimental knockout and knockin rodent strains. Mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and are grown under culture conditions well known in the art. Vectors for knockout strains contain a disease gene candidate modified to include a marker gene which disrupts transcription and/or translation *in vivo*. The vector is introduced into ES cells by transformation methods such as electroporation, adenoviral delivery, microinjection, and the like which are well known in the art. The endogenous rodent gene is replaced by the disrupted disease gene through homologous recombination and integration during cell division. Then transformed ES cells are selected under conditions, identified, and preferably microinjected into mouse cell blastocyst such as those from the C57BL/6 mouse strain. The blastocyst are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains.

ES cells are also used to study the differentiation of various cell types and tissues *in vitro*, such as neural cells, hematopoietic lineages, and cardiomyocytes (Bain *et al.* (1995) *Dev Biol* 168:342-357; Wiles and Keeler (1991) *Development* 111:259-267; and Klug *et al.* (1996) *J Clin Invest* 98:216-224). Recent developments demonstrate that ES cells derived from human blastocyst may also be manipulated *in vitro* to differentiate into eight separate cell lineages, including endoderm, mesoderm, and ectodermal cell types (Thomson (1998) *Science* 282:1145-1147).

Knockout Analysis

WO 02/04513

PCT/US01/21982

In gene knockout analysis, a region of a human disease gene candidate is enzymatically modified to include a non-mammalian gene such as the neomycin phosphotransferase gene (*neo*; Capecchi (1989) Science 244:1288-1292). The inserted coding sequence disrupts transcription and translation of the targeted gene and prevents biochemical synthesis of the disease candidate protein.

5 The modified gene is transformed into cultured embryonic stem cells (described above), the transformed cells are injected into rodent blastulae, and the blastulae are implanted into pseudopregnant dams. Transgenic progeny are crossbred to obtain homozygous inbred lines.

Knockin Analysis

Totipotent ES cells, present in the early stages of embryonic development, can be used to
10 create knockin humanized animals (pigs) or transgenic animal models (mice or rats) of human diseases. With knockin technology, a region of a human gene is injected into animal ES cells, and the human sequence integrates into the animal cell genome by recombination. Totipotent ES cells which contain the integrated human gene are handled as described above. Inbred animals are studied and treated to obtain information on the analogous human condition. These methods have been used
15 to model several human diseases. (See, e.g., Lee *et al.* (1998) Proc Natl Acad Sci 95:11371-11376; Baudoin *et al.* (1998) Genes Dev 12:1202-1216; and Zhuang *et al.* (1998) Mol Cell Biol 18:3340-3349).

Non-Human Primate Model

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

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

35

EXAMPLES

WO 02/04513

PCT/US01/21982

It is to be understood that this invention is not limited to the particular machines, materials and methods described. Although particular embodiments are described, equivalent embodiments may be used to practice the invention. The described embodiments are not intended to limit the scope of the invention which is limited only by the appended claims. The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention. For purposes of example, the preparation of the human hippocampal cDNA library, HIPONON02, is described.

I cDNA Library Construction

The human hippocampal cDNA library, HIPONON02, was constructed from tissue obtained from the hippocampus of a 72-year-old Caucasian female who died from an intracranial hemorrhage. The frozen tissue was homogenized and lysed in TRIZOL reagent (1 g tissue/10 ml; Life Technologies), using a POLYTRON homogenizer (PT-3000; Brinkmann Instruments Westbury NY). Following homogenization, chloroform was added (1:5 v/v chloroform:homogenate), and the lysate was centrifuged. The aqueous layer was removed, and the RNA was precipitated with isopropanol. The RNA was resuspended in DEPC-treated water and digested with DNase I (Life Technologies) for 25 min at 37C. The RNA was re-extracted with acid phenol-chloroform, pH 4.7, and precipitated using 0.3M sodium acetate and 2.5 volumes ethanol.

Messenger RNA (mRNA) was isolated using the OLIGOTEX kit (Qiagen, Valencia CA) and used to construct the cDNA library. The mRNA was handled according to the recommended protocols in the SUPERScript plasmid system (Life Technologies) which contains a NotI primer-adaptor designed to prime the first strand cDNA synthesis at the poly(A) tail of mRNAs. Double stranded cDNA was blunted, ligated to EcoRI adaptors, and digested with NotI (New England Biolabs, Beverly MA). The cDNAs were fractionated on a SEPHAROSE CL-2B column (APB), and those cDNAs exceeding 800 bp were ligated into the NotI and EcoRI sites of the PSPORT1 plasmid (Life Technologies). The plasmid was transformed into competent DH5 α , DH12s cells or DH10B cells (all Life Technologies).

II Normalization of cDNA Libraries

For purposes of example, the normalization of a human brain library is described. About 1.13×10^6 independent clones of the HIPONOT01 plasmid library in *E. coli* strain DH12S competent cells (Life Technologies) were grown in liquid culture under carbenicillin (25 mg/l) and methicillin (1 mg/ml) selection following transformation by electroporation. To reduce the number of excess cDNA copies according to their abundance levels in the library, the cDNA library was then normalized in a single round according to the procedure of Soares *et al.* (1994, Proc Natl Acad Sci 91:9228-9232), with the following modifications. The primer to template ratio in the primer extension reaction was increased from 2:1 to 10:1. The ddNTP concentration in the reaction was reduced to 150 μ M for each ddNTP to allow the generation of longer (400-1000 bases) primer

WO 02/04513

PCT/US01/21982

extension products. The reannealing hybridization was extended from 13 to 48 hr. The single stranded DNA circles of the normalized library were purified by hydroxyapatite chromatography, converted to partially double-stranded by random priming, and electroporated into *E. coli* strain DH10B competent cells (Life Technologies).

5 III Construction of pINCY Plasmid

The plasmid was constructed by digesting the PSPORT1 plasmid (Life Technologies) with EcoRI restriction enzyme (New England Biolabs, Beverly MA) and filling the overhanging ends using Klenow enzyme (New England Biolabs) and 2'-deoxymucleotide 5'-triphosphates (dNTPs). The plasmid was self-ligated and transformed into the bacterial host, *E. coli* strain JM109.

10 An intermediate plasmid produced by the bacteria (PSPORT1-ARD), which showed no digestion with EcoRI, was digested with Hind III (New England Biolabs); and the overhanging ends were again filled in with Klenow and dNTPs. A linker sequence was phosphorylated, ligated onto the 5' blunt end, digested with EcoRI, and self-ligated. Following transformation into JM109 host cells, plasmids were isolated and tested for preferential digestibility with EcoRI, but not with Hind III. A
15 single colony that met this criteria was designated pINCY plasmid.

After testing the plasmid for its ability to incorporate cDNAs from a library prepared using NotI and EcoRI restriction enzymes, several clones were sequenced; and a single clone containing an insert of approximately 0.8 kb was selected from which to prepare a large quantity of the plasmid.

After digestion with NotI and EcoRI, the plasmid was isolated on an agarose gel and purified using a

20 QIAQUICK column (Qiagen) for use in library construction.

IV Isolation and Sequencing of cDNA Clones

Plasmid DNA was released from the cells and purified using either the MINIPREP kit (Edge Biosystems, Gaithersburg MD) or the REAL PREP 96 plasmid kit (Qiagen). Such kits consist of a 96-well block with reagents for 960 purifications. The recommended protocol was employed except for

25 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) after incubation for 19 hours, the cells were lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were transferred to a 96-well block for storage at 4C.

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

WO 02/04513

PCT/US01/21982

with solution volumes of 0.25x-1.0x concentrations. In the alternative, cDNAs were sequenced using solutions and dyes from APB.

V Extension of cDNA Sequences

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

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

High fidelity amplification was obtained by PCR using methods such as that taught in USPN 5,932,451. PCR was performed in 96-well plates using the DNA ENGINE thermal cycler (MJ Research). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer (containing Mg²⁺, (NH₄)₂SO₄, and β-mercaptoethanol), Taq DNA polymerase (APB), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B (Incyte Genomics): Step 1: 94C, three min; Step 2: 94C, 15 sec; Step 3: 60C, one min; Step 4: 68C, two min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68C, five min; Step 7: storage at 4C. In the alternative, the parameters for primer pair T7 and SK+ (Stratagene) were as follows: Step 1: 94C, three min; Step 2: 94C, 15 sec; Step 3: 57C, one min; Step 4: 68C, two min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68C, five min; Step 7: storage at 4C.

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

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

WO 02/04513

PCT/US01/21982

prior to religation into pUC18 vector (APB). For shotgun sequences, the digested nucleotide sequences were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and the agar was digested with AGARACE enzyme (Promega). Extended clones were religated using T4 DNA ligase (New England Biolabs) into pUC18 vector (APB), treated with Pfu DNA polymerase

5 (Stratagene) to fill-in restriction site overhangs, and transfected into *E. coli* competent cells.

Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37C in 384-well plates in LB/2x carbenicillin liquid media.

The cells were lysed, and DNA was amplified using primers, Taq DNA polymerase (APB) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94C, three min; Step 2: 94C, 10 15 sec; Step 3: 60C, one min; Step 4: 72C, two min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72C, five min; Step 7: storage at 4C. DNA was quantified using PICOGREEN quantitative reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the conditions described above. Samples were diluted with 20% dimethylsulfoxide (DMSO; 1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT cycle

15 sequencing kit (APB) or the ABI PRISM BIGDYE terminator cycle sequencing kit (Applied Biosystems).

VI Homology Searching of cDNAs 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 20 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 25 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 30 nucleotides and 10^{-15} for peptides. Homology was also evaluated by product score calculated as follows: the % nucleotide or amino acid identity [between the query and reference sequences] in BLAST is multiplied by the % maximum possible BLAST score [based on the lengths of query and reference sequences] and then divided by 100. In comparison with hybridization procedures used in the laboratory, the electronic stringency for an exact match was set at 70, and the conservative lower

WO 02/04513

PCT/US01/21982

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
5 direct pairwise comparison of either nucleic or amino acid molecules. BLAST programs are commonly used with gap and other parameters set to default settings, e.g.: Matrix: BLOSUM62; Reward for match: 1; Penalty for mismatch: -2; Open Gap: 5 and Extension Gap: 2 penalties; Gap x drop-off: 50; Expect: 10; Word Size: 11; and Filter: on. Identity is measured over the entire length of a sequence or some smaller portion thereof. Brenner *et al.* (1998; Proc Natl Acad Sci 95:6073-6078,
10 incorporated herein by reference) analyzed the BLAST for its ability to identify structural homologs by sequence identity and found 30% identity is a reliable threshold for sequence alignments of at least 150 residues and 40%, for alignments of at least 70 residues.

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

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

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

WO 02/04513

PCT/US01/21982

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 $\leq 1 \times 10^{-3}$. The templates were also subjected to frameshift FASTx against GENPEPT, and homolog match was defined as having an E-value of $\leq 1 \times 10^{-3}$.

5 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
10 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 nucleic acid molecule was further analyzed using MACDNASIS PRO software (Hitachi Software Engineering), and LASERGENE software (DNASTAR) and queried against public
15 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
20 Génethon are used to determine if any of the cDNAs presented in the Sequence Listing have been mapped. Any of the fragments of the nucleic acid molecule encoding DSCR1L1 α that have been mapped result in the assignment of all related regulatory and coding sequences to the same location. The genetic map locations are described as ranges, or intervals, of human chromosomes. The map position of an interval, in cM (which is roughly equivalent to 1 megabase of human DNA), is measured
25 relative to the terminus of the chromosomal p-arm.

VII Hybridization Technologies and Analyses

Immobilization of cDNAs on a Substrate

The cDNAs are applied to a substrate by one of the following methods. A mixture of cDNAs is fractionated by gel electrophoresis and transferred to a nylon membrane by capillary transfer.

30 Alternatively, the cDNAs are individually ligated to a vector and inserted into bacterial host cells to form a library. The cDNAs are then arranged on a substrate by one of the following methods. In the first method, bacterial cells containing individual clones are robotically picked and arranged on a nylon membrane. The membrane is placed on LB agar containing selective agent (carbenicillin, kanamycin, ampicillin, or chloramphenicol depending on the vector used) and incubated at 37C for 16 hr. The

WO 02/04513

PCT/US01/21982

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 µg. Amplified nucleic acids from about 400 bp to about 5000 bp in length are purified using SEPHACRYL-400 beads (APB). Purified nucleic acids are arranged on a nylon membrane manually or using a dot/slot blotting
10 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) by ultrasound in 0.1% SDS and acetone, etching in 4% hydrofluoric acid (VWR Scientific Products, West Chester PA), coating with 0.05% aminopropyl
15 silane (Sigma-Aldrich, St Louis MO) in 95% ethanol, and curing in a 110C oven. The slides are washed extensively with distilled water between and after treatments. The nucleic acids are arranged on the slide and then immobilized by exposing the array to UV irradiation using a STRATALINKER UV-crosslinker (Stratagene). Arrays are then washed at room temperature in 0.2% SDS and rinsed three times in distilled water. Non-specific binding sites are blocked by incubation of arrays in 0.2%
20 casein in phosphate buffered saline (Tropix, Bedford MA) for 30 min at 60C; then the arrays are washed in 0.2% SDS and rinsed in distilled water as before.

Probe Preparation for Membrane Hybridization

Hybridization probes derived from the cDNAs of the Sequence Listing are employed for screening cDNAs, mRNAs, or genomic DNA in membrane-based hybridizations. Probes are prepared
25 by diluting the cDNAs to a concentration of 40-50 ng in 45 µl 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 µl 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 µl 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

WO 02/04513

PCT/US01/21982

cDNAs of the Sequence Listing in array-based hybridizations. Probe is prepared using the GEMbright kit (Incyte Genomics) by diluting mRNA to a concentration of 200 ng in 9 μ l TE buffer and adding 5 μ l 5x buffer, 1 μ l 0.1 M DTT, 3 μ l Cy3 or Cy5 labeling mix, 1 μ l RNase inhibitor, 1 μ l reverse transcriptase, and 5 μ l 1x yeast control mRNAs. Yeast control mRNAs are synthesized by *in vitro* transcription from noncoding yeast genomic DNA (W. Lei, unpublished). As quantitative controls, one set of control mRNAs at 0.002 ng, 0.02 ng, 0.2 ng, and 2 ng are diluted into reverse transcription reaction mixture at ratios of 1:100,000, 1:10,000, 1:1,000, and 1:100 (w/w) to sample mRNA respectively. To examine mRNA differential expression patterns, a second set of control mRNAs are diluted into reverse transcription reaction mixture at ratios of 1:3, 3:1, 1:10, 10:1, 1:25, and 25:1 (w/w). The reaction mixture is mixed and incubated at 37C for two hr. The reaction mixture is then incubated for 20 min at 85C, and probes are purified using two successive CHROMA SPIN+TE 30 columns (Clontech, Palo Alto CA). Purified probe is ethanol precipitated by diluting probe to 90 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 Na_2HPO_4 , 5 mM EDTA, pH 7) at 55C for two hr. The probe, diluted in 15 ml fresh hybridization solution, is then added to the membrane. The membrane is hybridized with the probe at 55C for 16 hr. Following hybridization, the membrane is washed for 15 min at 25C in 1mM Tris (pH 8.0), 1% Sarkosyl, and four times for 15 min each at 25C in 1mM Tris (pH 8.0). To detect hybridization complexes, XOMAT-AR film (Eastman Kodak, Rochester NY) is exposed to the membrane overnight at -70C, developed, and examined visually.

Polymer Coated Slide-based Hybridization

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

WO 02/04513

PCT/US01/21982

absolute hybridization format, probe from one sample is hybridized to array elements, and signals are detected after hybridization complexes form. Signal strength correlates with probe mRNA levels in the sample. In the differential hybridization format, differential expression of a set of genes in two biological samples is analyzed. Probes from the two samples are prepared and labeled with different
5 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 (Shaton WO95/35505).

Hybridization complexes are detected with a microscope equipped with an Innova 70 mixed
10 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
15 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
20 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.
25 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
30 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).

VIII Electronic Analysis

WO 02/04513

PCT/US01/21982

BLAST was used to search for identical or related molecules in the GenBank or LIFESEQ databases (Incyte Genomics). The product score for human and rat sequences was calculated as follows: the BLAST score is multiplied by the % nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences), such that a 100% alignment over the length of the shorter sequence gives a product score of 100. The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and with a product score of at least 70, the match will be exact. Similar or related molecules are usually identified by selecting those which show product scores between 8 and 40.

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

IX Complementary Molecules

Molecules complementary to the nucleic acid molecule, from about 5 (PNA) to about 5000 bp (complement of a cDNA insert), are used to detect or inhibit gene expression. These molecules are selected using OLIGO primer analysis software (Molecular Biology Insights). Detection is described in Example VII. To inhibit transcription by preventing promoter binding, the complementary molecule is designed to bind to the most unique 5' sequence and includes nucleotides of the 5' UTR upstream of the initiation codon of the open reading frame. Complementary molecules include genomic sequences (such as enhancers or introns) and are used in "triple helix" base pairing to compromise the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. To inhibit translation, a complementary molecule is designed to prevent ribosomal binding to the mRNA encoding the mammalian protein.

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

WO 02/04513

PCT/US01/21982

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 nucleic acid molecule encoding the mammalian protein.

X Expression of DSCRIL1 α

Expression and purification of the mammalian protein are achieved using either a mammalian cell expression system or an insect cell expression system. The pUB6/V5-His vector system (Invitrogen, Carlsbad CA) is used to express DSCRIL1 α in CHO cells. The vector contains the selectable bsd gene, multiple cloning sites, the promoter/enhancer sequence from the human ubiquitin C gene, a C-terminal V5 epitope for antibody detection with anti-V5 antibodies, and a C-terminal polyhistidine (6xHis) sequence for rapid purification on PROBOND resin (Invitrogen). Transformed cells are selected on media containing blasticidin.

Spodoptera frugiperda (Sf9) insect cells are infected with recombinant *Autographica californica* nuclear polyhedrosis virus (baculovirus). The polyhedrin gene is replaced with the mammalian nucleic acid molecule by homologous recombination and the polyhedrin promoter drives 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

DSCRIL1 α is purified using polyacrylamide gel electrophoresis and used to immunize mice or rabbits. Antibodies are produced using the protocols below. Alternatively, the amino acid sequence of DSCRIL1 α is analyzed using LASERGENE software (DNASTAR) to determine regions of high antigenicity. An antigenic epitope, usually found near the C-terminus or in a hydrophilic region is selected, synthesized, and used to raise antibodies. Typically, epitopes of about 15 residues in length are produced using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc-chemistry and coupled to KLH (Sigma-Aldrich) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester to increase antigenicity.

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

WO 02/04513

PCT/US01/21982

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 Nucleic Acid Molecule or Protein

The nucleic acid molecule or the protein are labeled with ^{32}P -dCTP, Cy3-dCTP, or Cy5-dCTP (APB), or with BIODIPY or FITC (Molecular Probes, Eugene OR), respectively. Libraries of candidate molecules or compounds previously arranged on a substrate are incubated in the presence of labeled nucleic acid molecule 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 mammalian protein of the invention. A cDNA encoding the protein is inserted into the multiple cloning site of a pLexA vector, ligated, and transformed into *E. coli*. cDNA, prepared from mRNA, is inserted into the multiple cloning site of a pB42AD vector, ligated, and transformed into *E. coli* to construct a cDNA library. The pLexA plasmid and pB42AD-cDNA library constructs are isolated from *E. coli* and used in a 2:1 ratio to co-transform competent yeast EGY48[p8op-lacZ] cells using a polyethylene glycol/lithium acetate protocol. Transformed yeast cells are plated on synthetic dropout (SD) media lacking histidine (-His), tryptophan (-Trp), and uracil (-Ura), and incubated at 30C until the colonies have grown up and are counted. The colonies are pooled in a minimal 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

WO 02/04513

PCT/US01/21982

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.

- 5 Histidine-requiring colonies are grown on SD/Gal/Raf/X-Gal/-Trp/-Ura, and white colonies are isolated and propagated. The pB42AD-cDNA plasmid, which contains a cDNA encoding a protein that physically interacts with the mammalian protein, is isolated from the yeast cells and characterized.

XIV Demonstration of Protein Activity

- DSCRIL1 α , or biologically active fragments thereof, are labeled with ¹²⁵I Bolton-Hunter reagent (Bolton *et al.* (1973) *Biochem J* 133:529-539). Candidate ligand molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled DSCRIL1 α , washed, and any wells with labeled DSCRIL1 α complex are assayed. Data obtained using different concentrations of DSCRIL1 α are used to calculate values for the number, affinity, and association of DSCRIL1 α with the candidate ligand molecules.

15

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

WO 02/04513

PCT/US01/21982

TABLE 1

Nucleic Acid SEQ ID NO:	Incyte Sequence Identifier	Nucleotide Length	Organism	Overlap with SEQ ID NO:1	% Identity with SEQ ID NO:1
3	219435.1	1021	Rat	443-954 1040-1212	86 87
4	271176.1	193	Rat	1323-1461	91
5	291862.1	1348	Rat	2419-2517 2158-2299	94 86
6	299263.1	531	Rat	668-794 884-954	86 89
7	219652.1	233	Rat	1683-1765 1593-1644	89 90
8	206751.1	615	Rat	734-837	82
9	702110167H1	297	Monkey	218-274	94

WO 02/04513

PCT/US01/21982

What is claimed is:

1. An isolated nucleic acid molecule or a fragment thereof encoding a protein or a portion thereof selected from:
 - 5 a) an amino acid sequence of SEQ ID NO:2;
 - b) an antigenic epitope of SEQ ID NO:2; and
 - c) a variant having at least 95% identity to the amino acid sequence of SEQ ID NO:2.
2. An isolated nucleic acid molecule or the complement thereof selected from:
 - a) a nucleic acid sequence of SEQ ID NO:1;
 - 10 b) a cDNA having at least 85% identity to the nucleic acid sequence of SEQ ID NO:1; and
 - c) a fragment of SEQ ID NO:1 from about nucleotide 1 to about nucleotide 400 of SEQ ID NO:1.
3. The cDNA of claim 2 selected from SEQ ID NOs:3-9.
- 15 4. A composition comprising the nucleic acid molecule or the complement of the nucleic acid molecule of claim 1 and a reporter molecule.
5. A substrate comprising the nucleic acid molecule or the complement of the nucleic acid molecule of claim 1.
6. A n array containing the nucleic acid molecule or the complement of the nucleic acid molecule of claim 1.
- 20 7. A vector comprising the nucleic acid molecule of claim 1.
8. A host cell comprising the vector of claim 7.
9. A method for producing a protein, the method comprising:
 - a) culturing the host cell of claim 8 under conditions for protein expression; and
 - 25 b) recovering the protein from the host cell culture.
10. A transgenic cell line or organism comprising the vector of claim 7.
11. A method for detecting differential expression of a nucleic acid molecule in a sample containing nucleic acids comprising:
 - a) hybridizing the nucleic acid molecule of claim 1 to nucleic acids in the sample, thereby
 - 30 forming hybridization complexes; and
 - b) quantifying the hybridization complexes, wherein the presence of hybridization complexes indicates expression of the nucleic acid molecule in the sample.
12. The method of claim 11 further comprising amplifying the nucleic acids of the sample prior to hybridization.

WO 02/04513

PCT/US01/21982

13. The method of claim 11 wherein detection of differential expression is diagnostic of Alzheimer's disease, Down syndrome or other forms of dementia.
14. A method of using a nucleic acid molecule to screen a plurality of molecules or compounds, the method comprising:
- 5 a) combining the nucleic acid molecule of claim 1 with a plurality of molecules or compounds under conditions to allow binding; and
- b) detecting binding, thereby identifying a molecule or compound which binds the nucleic acid molecule.
15. The method of claim 14 wherein the molecules or compounds are selected from DNA
- 10 molecules, RNA molecules, peptide nucleic acids, artificial chromosome constructions, peptides, transcription factors, repressors, and regulatory molecules.
16. A purified protein or a portion thereof selected from:
- a) an amino acid sequence of SEQ ID NO:2;
- b) an antigenic epitope of SEQ ID NO:2;
- 15 c) a variant having at least 95% identity to the amino acid sequence of SEQ ID NO:2.
17. A composition comprising the protein of claim 16 and a reporter molecule.
18. A method for using a protein to screen a plurality of molecules or compounds to identify at least one ligand, the method comprising:
- a) combining the protein of claim 16 with the molecules or compounds under conditions
- 20 to allow specific binding; and
- b) detecting specific binding, thereby identifying a ligand which specifically binds the protein.
19. A method of using a protein to prepare and purify antibodies comprising:
- a) immunizing a animal with the protein of claim 16 thereby eliciting an antibody response;
- 25 b) isolating animal antibodies;
- c) attaching the protein to a substrate;
- d) contacting the substrate with isolated antibodies under conditions to allow specific binding to the protein;
- e) dissociating the antibodies from the protein, thereby obtaining purified antibodies which
- 30 specifically bind the protein.
- 20 An antibody which binds the protein of claim 16.

```

190      199      208      217
5'  AUG  TCA  ACT  TTG  GAA  AAA  AAC  AAC  GTA  GTG  TNG
-----
      M  S  T  L  E  K  N  N  V  V  X
226      235      244      253      262      271
GGA  ATG  AGG  GGA  GAA  TCA  TAC  TTC  ATC  GGA  ATG  AGG  AGC  CCA  GGG  CAG  CAG  GGA
-----
G  M  R  G  E  S  Y  F  I  G  M  R  S  P  G  Q  Q  G
280      289      298      307      316      325
CAC  GTC  CCT  GAA  GAT  GGA  GGA  CTT  TTC  TTA  CTG  TGC  TGC  ATA  GAC  AGG  GAC  TGG
-----
H  V  P  E  D  G  G  L  F  L  C  C  I  D  R  D  W
334      343      352      361      370      379
GCT  GTC  ACT  CGT  TGT  TTT  GCA  GAA  GAA  GCC  TTT  CAA  GCA  ATC  ACT  GAC  TTC  AAT
-----
A  V  T  R  C  F  A  E  E  A  F  Q  A  I  T  D  F  N
388      397      406      415      424      433
GAC  CTC  CCC  AAC  TCG  TTG  TTT  GCG  TGC  AAT  GTT  CAC  CAG  TCA  GTG  TTT  GAA  GGA
-----
D  L  P  N  S  L  F  A  C  N  V  H  Q  S  V  F  E  G
442      451      460      469      478      487
GAA  GAG  AGC  AAG  GAA  AAA  TTT  GAG  GGA  CTG  TTT  CGG  ACT  TAT  GAT  GAC  TGT  GTG
-----
E  E  S  K  E  K  E  G  L  F  R  T  Y  D  D  C  V

```

FIGURE 1A

```

496  AC G TTC CAG CTA TTT AAG AGT TTC AGA CGT GTC CGT ATA AAC TTC AGC AAT CCT 541
-----
505  T F Q L F K S F R R V R I N F S N P
-----
514  AAA TCT GCA GCC CGA GCT AGG ATA GAG CTT CAT GAA ACC CAA TTC AGA GGG AAA 595
-----
559  K S A A A R A R I E L H E T Q F R G K
-----
613  AAA TTA AAG CTC TAC TTT GCA CAG GTT CAG ACT CCA GAG ACA GAT GGA GAC AAA 649
-----
622  K L K L Y F A Q V Q T P E T D G D K
-----
667  CTG CAC TTG GCT CCA CCC CAG CCT GCC AAA CAG TTT CTC ATC TCG CCC CCT TCC 703
-----
721  L H L A P P Q P A K Q F L I S P P S
-----
730  TCC CCA CCT GTT GGC TGG CAG CCC ATC AAC GAT GCC ACG CCA GTC CTC AAC TAT 757
-----
748  S P P V G W Q P I N D A T P V L N Y
-----
775  GAC CTC CTC TAT GCT GTG GCC AAA CTA GGA CCA GGA GAG AAG TAT GAG CTC CAT 811
-----
784  D L L Y A V A K L G P G E K Y E L H
-----

```

FIGURE 1B

WO 02/04513

3/5

PCT/US01/21982

```

820  CCA GGG ACT GAG TCC ACC CCA AGT GTC GTC GTG CAC GTG TGC GAC AGT GAC ATA 865
-----
829  A G T E S T P S V V V H V C D S D I 866
-----
874  GAG GAA GAA GAG GAC CCA AAG ACT TCC CCA AAG CCA AAA ATC AAA ACT CGG 919
-----
883  E E E E D P K T S P K P K I I Q T R
-----
928  CGT CCT GGC CTG CCA CCC TCC GTG TCC AAC 3'
-----
937  R P G L P P S V S N
-----
946

```

FIGURE 1C

149 P E T D G D K L H L A P P Q P A K Q F L I S P P S S P P V G D S C R I L L α
86 P E T D G D K L H L A P P Q P A K Q F L I S P P S S P P V S g1435040
127 S G E V R D K S Y L L P P Q F V K Q F L I S P P A S P P V G g6017919
179 W Q P I N D A T P V L N Y D L L Y A V A K L G P G E K Y E L D S C R I L L α
116 W Q P I N D A T P V L N Y D L L Y A V A K L G P G E K Y E L g1435040
157 W K Q S E D A M P V I N Y D L L C A V S K L G P G E K Y E L g6017919
209 H A G T E S T P S V V V H V C D S D I E E E E D P K T S P K D S C R I L L α
146 H A G T E S T P S V V V H V C D S D I E E E E D P K T S P K g1435040
187 H A G T E S T P S V V V H V C E S E T E E E E E T K - N P K g6017919
239 P K I I Q T R R R P G L P P S V S N D S C R I L L α
176 P K I I Q T R R R P G L P P S V S N g1435040
216 Q K I A Q T R R R P D P P T A A L N E P Q T F D C A L g6017919

FIGURE 2B

WO 02/04513

PCT/US01/21982

<110> INCYTE GENOMICS, INC.
LORING, Jeanne F.
TINGLEY, Debora W.
EDWARDS, Carla M.
STREBETER, David G.

<120> DOWN SYNDROME CRITICAL REGION 1-LIKE 1 PROTEINS

<130> PC-0013 PCT

<140> To Be Assigned

<141> Herewith

<150> 09/614,474

<151> 2000-07-11

<160> 13

<170> PERL Program

<210> 1

<211> 3159

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 247500.5

<220>

<221> unsure

<222> 216, 3091, 3093, 3103

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

<400> 1

```

ccagctatta gggagactcc agccccttgc caggcgagag agtggatggt caccctccat 60
ggaggaagtg tttccaagtg gtgtctgctg ggaaggaaa gcatgatgca gtgcagatta 120
atcagagaga agagccaacg tctctctctc ctttttttgt tgaaaacaaa caaaactgta 180
ttgtatgtca actttggaaa aaaacaacgt agtgnngga atgaggggag aatcataact 240
catcggaatg aggagccagg ggcagcagg acactgctct gaagatggag gacttttctt 300
actgtgctgc atagacaggg actgggctgt cactgttgtt ttgcagaag aagcctttca 360
agcaatcact gacttcaatg acctcccaca cctgtgtttt gcgtgcaatg ttcaccogtc 420
agtgtttgaa ggaagaagga gcaaggaaaa atttgaggga ctgtttcoga cttatgatga 480
ctgtgtgacg ttccagctat ttaagagttt cagacgtgtc cgtataact tcagcaatcc 540
taaactctga gccccagcta ggaatagact tcatgaaacc caattcagag ggaaaaaatt 600
aaagtcttac ttgcaacagg ttccagctcc agagacagat ggagacaaac tgcacttggc 660
tccaccaccg cctgccaacg agttctctat ctgcacctct tctcccacc ctgttggctg 720
gcagcccatc aacgatgcca gcaccgtcct caactatgac ctctctatg ctgtggccaa 780
actaggacca ggagagaagt atgagctcca tgcagggact gactccacc caagtgtctg 840
cgtgcacgtg tgcgacagtg acatagagga agaagaggac ccaaaagact ccccaagcc 900
aaaaatcatc caaactcggc gtctctggct gccaccctcc gtgtccaact gagctgctg 960
ctctctctcg ataatagccc tctctctctt atcatgctt ttcccctgt tgtttgtcaa 1020
aaaaaattgc ctttaaatc ctgggtgttt ggttgttga gatctctcc ttgtatcaa 1080
gcctctcga caaaaggctc aggaaaaggt gatatgtctc ctgatcatat cataccatt 1140
aagtataacc cattatttag aaggttctag ggaaaaaagt agtattttct tattaacaa 1200
tcagcacagc ctatatcttt gttctctcat gttgatccaa gccagagaca tcagtaacaa 1260

```

WO 02/04513

PCT/US01/21982

```

atagcacctg tgttgtttgt gagctgtttc agtcccagtc ctgatgtgtg tgcgttgttc 1320
tctccgggcc acttaaatag gaccatagt aaacttgact ttgactgcat gagatatecc 1380
tatctggctc cactcagtcct tctgctcccc aaacttccca ggacatgcat gatcaccagc 1440
ttttattttc attatttgag gatcacttat aactcacaga ttgtcagcat ccagccatgt 1500
cctatctaga ttaggaaaat gatcagaata ttccagctca acaagtctgg gtatattcac 1560
tattgtgagt caatacacca tagctctggt gaaattctcg gaggcaaat tgacctggc 1620
cccaaagata ttctcaata gatttcaaac accactcccc tgtagaactc tcccagctcc 1680
gttggggagg cttgtccagc gtgatagaga ctgatttcag acaaacctat ttattacaaa 1740
agtttcatgg tgcctgaatg attgttttct ctctttgtat atttgtacaa atgtttcagc 1800
tgtgtcttta aaaaatctgg atgtttttta tttagtgtat ttctgacaa tagctgtctc 1860
aaaacataat gtgcattgct tatgaatgcc ttcatatact aatacagata ctctgataat 1920
attacactct aataaggata atgctgaatt ttgaaaggac acaaaacatc laatgccaat 1980
atatacatga ttaggcaaca tctttgctat caagaccact tgtttttaa taaagatgca 2040
agtgctagtt tagattatt gggatgagc taaatcccc gaatggcagc agcagctgag 2100
catgttaaaa tggggaagaa tgatagctac atgtatgccg gtccctactca cgcgaccccc 2160
gtgtgctcaa aaaagtatt tgtttttgtt acgtgtgtat tttctatttc tctagcccaa 2220
agtgcatcac agaagataca cctatagaac cattacotc tgcctatgtt gccaggctcc 2280
atctactcct gtacatbaat gaattacttt agatgcaaac cgagattaca atggagtggt 2340
gaagtacttt cattaccocaa gcctcagaaa aacacacaag aacaataaca cagcaaacag 2400
attgagggat tttgttgggt ttgactaag gtgtatgtta gtttcatcag aaacttaaaa 2460
catagactga tcactoagaa attaaagctc gttttactgt gaatatgaca atatagtact 2520
ggacscagta cbgtgaaac tggaggagac attgctgtga aaatcctgag ttccataag 2580
gaaaatgaaa actcctttta aaaaataaat ctgaggagtg tacaataagc atatgctttg 2640
acttctcttt cgtgtggagg tttttggttt ttcatgtgat ataaacgact acagacttag 2700
tagtggagaa atggtgtcct ctagtgaag aaatagtagg ctccgctatt cagatgcaga 2760
gcactgcagc atccagcctt tcaaacgtga ctctctocaa tcatctgtgg gtcatttgac 2820
ttgatttttt aagctaccct gaattccag aatgcagggt ctaaagaaat ctagatgaga 2880
gaaagtattt gaaaatgatt ttttaattgt ttttaaaaga cacatctgac atttttaaca 2940
acttagtaaa agttgaaatg accattctgt gtagtcataa aagaacacaa atgaagtgta 3000
tggcctctgg agttagtctt agttaaactt attgctctgt gtcaatgtta acctgtctca 3060
gatcaagtaa ttcttcaact aggttgggtt nongaggggg ganaagagggg gctttctctag 3120
gagaacgata agaaatggaa agactccttg aagtgttgc 3159

```

```

<210> 2
<211> 255
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 247500.5

<220>
<221> unsure
<222> 11
<223> unknown or other

<400> 2
Met Ser Thr Leu Glu Lys Asn Asn Val Val Xaa Gly Met Arg Gly
1 5 10 15
Glu Ser Tyr Phe Ile Gly Met Arg Ser Pro Gly Gln Gln Gly His
20 25 30
Val Pro Glu Asp Gly Gly Leu Phe Leu Leu Cys Cys Ile Asp Arg
35 40 45
Asp Trp Ala Val Thr Arg Cys Phe Ala Glu Glu Ala Phe Gln Ala
50 55 60

```

WO 02/04513

PCT/US01/21982

```

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

```

```

<210> 3
<211> 1021
<212> DNA
<213> Rattus norvegicus

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 219435_Rn.1

```

```

<400> 3
aggctgtact gattgtttaa taagaaaata ctacccttgg tctagaactt tctaaataat 60
ggactctggg taatgggtat gagttgactg ggagacatat caccttttcc tagccctttt 120
gtccgagagg cctaaagttac aaggaaggat tctcaacaaa ccaaataccg aggtgacttc 180
aaggcaatta atttttggac gaacaaggga gaaaaagcat gataagagga gatgcatgtt 240
atcatggacg tgctcagttg gacacggagg gtgycaggcc aggcgcgggt gtcgggatga 300
tttttggctt tggggaagtc tttgatcctt cctcctctcc caagtgcctg tcacacagct 360
gcacgacaac gctcgggtga gactcagttc ccgcatgcag ctcatatttc tctcctggtc 420
ctagtttggc cacggcataa aggaggtcgt acgttgagga ctggtgtggc atcgctgata 480
ggcttccagc caacgggagg agatgaaggg ggtgagatga ggaactgttt ggcaggctgt 540
gggggtgcca aactgcagtt tgtctaccat ctgtctctgg ggtctggacc tgtgcaaat 600
agagtttccg cttcttccct catgaactgg gtccacatga agtctctatc cgggcaccgg 660
gctgcagctt tgggggtggc gaaatttatt ccgaaccogt cygaaactct taaacagctg 720
aaagtcaca cattcgtcat aggtccggaa cagtcctcca aatttttctt taacctcctg 780
attgtaaaag acctccacat ccaccacaca ggcgaccaga gtggaacat cacagtccat 840
gctaggggct ggcattcccc tcttcggaat tctcagcga gagttaggca gcctcagagt 900
tgggagtgaa gagactcttc tggatggata ctgtcctggc actgccctgc tctcgggggg 960
gaggtggggt gctgtgaggt gctggttggg gactcagcca cctggagcct gctgcccctg 1020
c 1021

```

WO 02/04513

PCT/US01/21982

<210> 4
<211> 193
<212> DNA
<213> Rattus norvegicus

<220>
<221> misc_feature
<223> Incyte ID No: 271176_Rn.1

<400> 4
tcttggccac ttatatagga ccataagtac acttggcttt gactgcatga gatcccccta 60
tctgtctcca cccagtccctc tgcattcccga cattcccagg acatgcacaa tcaccagcat 120
ttattttgat gatttgaggg tatgcccacc ctigtgtgat cccagccctg tccgttatag 180
acaaggggca tcc 193

<210> 5
<211> 1348
<212> DNA
<213> Rattus norvegicus

<220>
<221> misc_feature
<223> Incyte ID No: 291862_Rn.1

<400> 5
ggctacagca atgtctcatg ccaatagata aacgactagc taatttcttt gctgtcaaga 60
tcacctgttt gcaaacaaag atccaagtat cagttgtaag caacttgggt ggcgcttaat 120
ccaagatcgt atacagctga gcatttccga acacaaagga tgatgggtaa acggacggcc 180
cctcccattc atccagtgcc cgtgtgctca aaagtccott gttttcgtta cctgtgatit 240
tgtatttctc tagcccaaaag tgcattacag aagatacacc tgcagaacca tactttctgc 300
tccatgtgtc aagcctcatc tattcctgtt cattaactag taactttgga tgcaataact 360
attaccaca caacgggcaa ggactcaat gacctaaccc tccgaattaa ccaaacaaag 420
aacctaaca cagccacctg atgggggat tatgtggttt ttgactaag gtgtatgta 480
gtttcagcag aaacttcaaa cctagactga tcactcagaa aattaaagtc cgttctactg 540
tgaaatag caataatagtg gccggacacg gtacacgtac gcgttgagac cgaggccacg 600
attgcttgtg gaatcctgag ttccacggg gatacaagt ctcctttca aactcagatc 660
tgaggagtg acagcaaaag ctttgacttt cttttgtgg tggaggatt tggtttttc 720
attggccacg gaacgactac aaatagttgc gagatgctc cctctgttg cgaaacgct 780
gaactcggtt gtabaggtg attgattaca acagtcaact cccaggtct gactttctea 840
tccgcttaa ctttcaggat gcaggttctg agaagagag aaaaatgaac agtotaatg 900
agatlaaagc cttgaaata atttgtgctt ccaacaatg atagcagaga tttccagttt 960
agtlaaagtt gatgtgacta cctctgggt aggcattaa aactcagtg tacagatggc 1020
ccaggtgatt agtcttagta aactgtatt ccatatgtc atgttaacct gctgcagagc 1080
aaggattct cgcactaggt tgagcacgga gagggagaa gggagcctt cccagaaag 1140
aatagggaag ccatgggacc tcccagacc gagaagtcg cctctacat tccctccatg 1200
atgctcatta ctgcaaacct cttgtagcgc tegtttttaa gtctgcaaac gttttaaag 1260
gaggggaagg ggaaggtttc caccaactga atcatttgg cactgttaca gctcaaaag 1320
cttagagttc aaatataatc ggtgaatg 1348

<210> 6
<211> 531
<212> DNA
<213> Rattus norvegicus

<220>
<221> misc_feature

WO 02/04513

PCT/US01/21982

<223> Incyte ID No: 299263_Rn.1

```

<400> 6
gttgactggg gagacatac actcttatcc tagccotttt gtccgagagg cttaatttac 60
tatggaatgg attctcaaac aatccaataa ccaggtact tcaatggcca attaatTTTT 120
ggacgaacaa gggagaaaaa gcatgataag aggagatgca tgttatacat ggacgtgctc 180
agttggacac ggagaggtgg caggccagga cccgggtta ctggatagat ttataaggct 240
ttggggagag cttatggatc ctacctacct cctccaagtc gctgtcacac acgtagcacg 300
acaacgtcgg gtagtagact acagtatccc gcataacagc tcatatatta cactcctggg 360
cctagtttgg ccaacggata aaggaggtcg tagttgagga ctggtagtgg catcctgtgt 420
aggcttcagc caaacgggag gagatagaag ggggtagat gaggaactgt atggcacgct 480
gtggaggggtg ccaagatgca gtttagtctc catctgacta caggggtctg g 531

```

```

<210> 7
<211> 233
<212> DNA
<213> Rattus norvegicus

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 219652_Rn.1

```

```

<400> 7
aattcctgga ggtaaaattg agcgtgcccc caaagagatt cctcagtaga ttctaaacat 60
cactalcaca taggactacc cagtgctact ggggagctt gcccgagtga cagagacoga 120
tttcagacaa acctatttat tataaaagtc tcacagtgtc tgaatgattg tcytcccccc 180
ccaccctttt gtatatttgt acagatgttt gagctgtgct tttaaaatct gga 233

```

```

<210> 8
<211> 615
<212> DNA
<213> Rattus norvegicus

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 206751_Rn.1

```

```

<400> 8
gaagatgcta cgagacagcc tgaatcttgg tgaatgacag ccagtcagac ctctgtagca 60
ggcaccagga agatggaaga ggagatggtc ttoggtgaaa atgaagacga tttggaagag 120
atgatggacc tcagcgacct gcccaactcc ctctttgctt gcagtgtcca tgaacgggtg 180
tttgaggtcc aagagcagaa ggagaggttc gaagccctgt tcaacctcta tgatgaccag 240
gtcacattcc agctgttcaa gagctttcgc agagtgagga tcaacttcag caagcctgag 300
gctgcccoga gagcacggat cgagctccac gagagtgaat tccatggccc gaagctgaag 360
ctttacttcg cacaggtgca ggtgtccggg gagctcggg acaagtccta cttaetgcca 420
ccacagccca ccaacaggtt cctcatctcc cctcccgcct caccctcctg ggggtggaag 480
cagagtgaag atgccacgcc agtgatcaac tatgacctcc tctgtccgt ctccaagctg 540
gggcccaggg agaatacga actgcacgca ggaaccagat ccaccccagc tgtgctggtg 600
cagctctgtg agagc 615

```

```

<210> 9
<211> 297
<212> DNA
<213> Macaca fascicularis

```

<220>

WO 02/04513

PCT/US01/21982

<221> misc_feature

<223> Incyte ID No: 702110167H1

<400> 9

```

cagcagcccc cgggtcgtgc tgggggaccg cggcggggct ctggccgcgg gacgggtctg 60
ggtgcagggc gggggcggcg ggggggtgg ggtttgtctc cagatctgtg gtcagtcocg 120
gctgcggccg tggggacgga gatctcccgc gacccccctc tagggggcgc ggtcccaag 180
gagccccctc cccactgctc gtggcccggg ggttcgcctt cctggaagca gcagcaggaa 240
tgaggggaga atcatacttc atcggaatga ggagcccggy acagcagtga cacatcc 297

```

<210> 10

<211> 192

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: g1435040

<400> 10

```

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

```

<210> 11

<211> 241

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: g6017919

<400> 11

WO 02/04513

PCT/US01/21982

```

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

```

```

<210> 12
<211> 3184
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: g1435039

```

```

<400> 12
ctctgtgtgt ctgcctcaaa cgcgaggggc tgcgtgcagt gggagcgggc tccaggagcc 60
cgagcctcca gccgtctcca gagcaaggca gcaccgaggc ctggccacag caatatccat 120
ctggaagctc ttcccttcac tcccaactct gaggttgctt aactctttat taaaattca 180
gaaggggaaa tgccagcccc tagcatggac tgtgatgttt ccaactctgtt tgcctgtgtg 240
gtggatgtcg aggtctttac caatcaggag gtttaagaaa aatttggggg actgtttcgg 300
acttatgatg actgtgtgac gttccagcta ttttaagagt tcagacgtgt ccgtataaac 360
ttcagcaatc ctaaactctgc agcccagact aggatagagc ttcataaac ccaattcaga 420
gggaaaaaat taaagctcta ctttgacag gttcagactc cagagacaga tggagacaaa 480
ctgcaacttg ctcaccocca gcctgccaac cagttttctc tctgcccccc ttcctcccca 540
cctgttagct ggcagcccat caacgatgcc acgccagtcc tcaactatga cctcctctat 600
gctgtggcca aactaggacc aggagagaag tatgagctcc atgcagggac tgagtccacc 660
ccaagtgtcg tcgtgcacgt gtgcgacagt gacatagagg aagaagagga cccaaagact 720

```

WO 02/04513

PCT/US01/21982

```

tccccaaagc caaaaatcat ccaaactcgg cgtcctggcc tggccaccctc cgtgtccaac 780
tgagctgect gtcctctctc gataatagcc gtcctcctctt tatcatgctt tttcccctg 840
ttgtttgtca aaaaaaattg cctttaaatc cctgggtggt ttggttggtg agatcccttc 900
ctgtttatca agcctctcgg acaaaaagggc taggaaaagg tgatatgtct cctgatcata 960
tcatacccat taagtataac ccattattta gaaggttcta gggaaaaaag tagtatttcc 1020
ttattaacaa atcagcacag cctatatctt tgttctctca tggtagtcca agccagagac 1080
atcggtaaca aatagcaact gtgtgttttg tgaggtggtt cagtcccagt cctgatgtgt 1140
gtgctgtggt ctctcctggc caactaaata ggaccatag taaacttgac ttgactgca 1200
tgagatattc ctatctggtc tcactcagtc ctctgcattc caactctcc aggacatgca 1260
tgatccaccg catttatttt cattatttga ggatatttca taactcacag atgtcagca 1320
tcagccatg tcctatctag attagggaaa tgatcagaat attccagctc aacaagtctg 1380
cgtatgataa ctattgtgag tcaatacacc atagctctgt tgaattctc ggaggcaaaa 1440
ttgaccttgg ccccaaeagc attcctcaat agatttcaaa caccactccc ctgtagaact 1500
ctcccagcct cgttggggag gctgtccag ggtgatagag actgatattca gacaacctca 1560
ttattacaaa aagtttcatg gtgtctgaat gatgttttc tctcttcta tattgtaca 1620
aatgtttcag ctgtgctttt aaaaaactcg gatgtttttt attagtgat tgttcgaca 1680
ttagctgctt caaacataa tggctattgc ttatgaatgc cttcatatcc taatcacagt 1740
actctgataa tattacactc taataaggat aatgctgaat ttgaaagga cacaaaacct 1800
cfaatgccaa tatatacatg gttagccaac atctttgcta tcaagaccac ttgttttaa 1860
taaagatgca agtgcagtt gtagattatt gggatgaagc taaatcccca gaatgcagca 1920
gcagctgagc atgttaaaat ggggaaggat gatagctaca tgtatgccgg tctactcac 1980
gcagaccocg tgtgtcaca aagttactt gtttttgta cgtgtgattt tctctattct 2040
ctagcccaaa gtgcattaca gaagatacac ctatagaacc ataccctct cctatgtgtg 2100
ccagggctca tctactcctg tacatataat gattacttta gatgcaaat cagattacaa 2160
tggagtgggg aagtactttc attacccaag cctcagaaaa acacacaaga acaataacac 2220
agcaaacaga ttgagggatt gttgtggttt ttgactaagg tgatgttag ttcatcaga 2280
aacttaaac atagactgat cactcagaaa ttaaagtcg ttttactgt aatatagca 2340
tatagctagc gacacagtat tggtaaaact gaggagagca ttgctgtaa aatcctgagt 2400
ttccataaag aaaaagaaa ctcttttaa aataaaatc tgaggagtg acaataagca 2460
tatgtttga ctttctttg ctgtggaggt tttgtttt tcatgtaga taaacgacta 2520
cagacttagt agtgagaaa tgggtctctc tagtgaaga aatagtagt ccgctattca 2580
gatcgagagc actgcagcat ccagccttcc aaagctgact cttctcaatc atctgtgggt 2640
catttgactt gatttttaa gctaccctga atttcagaa tgcaggttct aaagaaatct 2700
agatgagaga aagtatttga aatgatattt taatgtttt ttaazagaca catctgacat 2760
ttttaacaac ttagtaaaag ttgaaatgac cattctgtgt agtcataaaa gaaacacaa 2820
gaagtgatg ccctctggag ttatcttag taaaactat tgcctctgt caatgtaac 2880
ctgtctcaga tcaagtatt ccttcaactag gttgggttg gggagggggg aaaagagggg 2940
cttttcttag gagaacgata agaaatggaa agactcctg aagtgttcca aggaaacctc 3000
ctagcactgt gaaagtcaga atgcctcag catttccatg acgcacatta tgcaaatctc 3060
tttagcacta ttttaaggtt gaaaacttta acaatgaagg ggaaggggaa gatctccacc 3120
aactgaaatc tttgtgcacg tgtatagctc aaagagctta gacttcaaat atatctggtg 3180
aatg 3184

```

```

<210> 13
<211> 828
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: g6017918

```

```

<400> 13
aaaaggccca ctttggggga taatgctgag ggacactatg aaatcttggg atgatagcca 60
gtcagatctg ttagoactg accaagaaga ggaagaagag atgatttttg gtgaaaatga 120
agatgatttg gatgagatga tggatttaag tgactgctc acctcacttt ttgcttgcag 180

```

WO 02/04513

PCT/US01/21982

```
cgtccatgaa gcagtgtttg aggcacgaga gcagaaggaa agatttgaag cactcttcaac 240
catctatgat gaccaggtta ctttcagct gtttaaaagc tttagaagag tcagaataaa 300
tttcagcaaa cctgaagcgg cagcaagagc gcgaatagaa ctccacgaaa cagacttcaa 360
tgggcagaaag ctaaagctat attttgaca ggtgcagatg tccggcgaag tgcgggacaa 420
gtcctatctc ctgcccctcc agcctgtcaa gcagttcctc atctcccctc cagcctctcc 480
cccagtgggg tggaaacaga gcgaagatgc gatgctgtt ataaattatg atttactctg 540
tgctgtttcc aaattgggac caggagagaa atatgaactt cacgcgggaa cagagtcgac 600
acccagcgtg gtggttcacg tctgtgaaag tgaactgaa gaggaagaag agacaaaaaa 660
ccccaaacag aaaattgcc agacaaggcg cccgaccct cgcaccgag cgttgaatga 720
gcccagacc tttgattgcg cgtgtgagg ccttggtg tggtgcgagg cggctgccct 780
ggtgggctct ggccatggcg ctctgtcct gcggccgatg cgttgctg 828
```

【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
17 January 2002 (17.01.2002)

PCT

(10) International Publication Number
WO 02/004513 A3

- (51) International Patent Classification: C12N 15/12, C07K 14/47, 16/18, C12Q 1/68, G01N 33/50, A01K 67/027
- (52) International Application Number: PCT/US01/21982
- (53) International Filing Date: 11 July 2001 (11.07.2001)
- (54) Filing Language: English
- (55) Publication Language: English
- (56) Priority Data: 09/614,474 11 July 2000 (11.07.2000) US
- (57) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (58) Inventors; and
(59) Inventors/Applicants (for US only): LORING, Jeanne, F. [US/US]; 701 Santa Cruz Lane, Foster City, CA 94044 (US). TINGLEY, Debora, W. [US/US]; 119 Knollview Way, San Francisco, CA 94131 (US). EDWARDS, Carla,
- (60) M. [US/US]; 41 Ventura Street, Half Moon Bay, CA 94019 (US). STREETER, David, G. [US/US]; 1211 Pinecrest Drive, Boulder Creek, CA 95006 (US).
- (61) Agents: HAMLET-COX, Diana et al.; Inceyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (62) Designated States (national): AF, AG, AI, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GE, GL, GM, GR, 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, MY, NZ, NI, NL, 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.
- (63) 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, GW, ML, MK, NE, NI, SN, TD, TG).
- (64) Published: with international search report

[Continued on next page]

(54) Title: DOWN SYNDROME CRITICAL REGION 1-LIKE 1 PROTEINS

```

          190       199       208       217
5'  ATG TCA ACT TTG GAA AAA AAC AAC GTA GTG TNG
    M  S  T  L  E  K  N  H  V  V  K
-----
226  235       244       253       262       271
GGA ATG AGG GGA GAA TCA TAC TTC ATC GGA ATG AGG AGC CCA GGG CAG CAG GGA
G  M  R  G  E  S  Y  F  I  G  N  R  S  P  G  Q  Q  G
-----
280  289       298       307       316       325
CAC GTC GCT GAA GAT GGA GGA CTT TTC TTA CTG TGC TGC ATA GAC AGG GAC TGG
H  V  P  E  D  G  G  L  F  L  L  C  C  I  D  R  D  W
-----
334  343       352       361       370       379
GCT GTC ACT CGT TGT TTT GCA GAA GAC TTT CAA GCA ATC ACU GAC TTC AAT
A  V  T  R  C  P  A  E  R  A  F  Q  A  I  T  D  F  N
-----
388  397       406       415       424       433
GAC CTC CCC AAC TCG TTG TTT GCG TGC AAT GTT CAC CAG TCA GGG TTT GAA GGA
D  L  P  N  S  L  P  A  C  H  V  H  Q  S  V  P  E  G
-----
442  451       460       469       478       487
GAA GAG AGC AAG GAA AAA TTT GAG GGA CTG TTT CCG ACT TAT GAT GAC TGT GCG
E  E  S  K  E  K  F  E  G  L  F  R  T  Y  D  P  C  V

```

(57) Abstract: The invention provides a nucleic acid molecule and fragments thereof. It also provides for the use of the nucleic acid molecule for the characterization, diagnosis, evaluation, and treatment of conditions, diseases and disorders associated with gene expression and for the production of a model system. The invention additionally provides expression vectors and host cells for the production of the protein encoded by the nucleic acid molecule.

WO 02/004513 A3

WO 02/004513 A3 

(88) **Date of publication of the international search 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.*
24 April 2003

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT		Info at Application No PCT/US 01/21982
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/12 C07K14/47 C07K16/18 C12Q1/68 G01N33/50 A01K67/027		
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 C12Q G01N A01K		
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)		
EMBL, BIOSIS, EPO-Internal, GENSEQ, WPI Data		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	WO 01 94629 A (AVALON PHARMACEUTICALS) 13 December 2001 (2001-12-13) SEQ ID NOS:105 and 3524 ---	1,2
L	DATABASE GENSEQ [Online] 15 May 2002 (2002-05-15) "Colon adenocarcinoma related gene sequence SEQ ID NO:105." Database accession no. ABL61768 XP002214753 cited in support of the sequences of W00194629 --- -/--	1,2
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document relating to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "Z" document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
8 October 2002		15. 01. 2003
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040. Tx. 31 651 epo nl. Fax: (+31-70) 340-3016		Authorized officer: MADDOX, A

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

INTERNATIONAL SEARCH REPORT		Internat'l Application No PCT/US 01/21982
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 GENESEQ [Online] 15 May 2002 (2002-05-15) "Lung cancer related gene sequence SEQ ID NO:3524." Database accession no. ABL65187 XP002214754 cited in support of the sequences of W00194629 ---	1,2
E	WO 02 04491 A (ROTHERMEL BEVERLY ;UNIV TEXAS (US); WILLIAMS SANDERS R (US)) 17 January 2002 (2002-01-17) SEQ ID NOS:17,18, and 19 ---	1,2,7,8, 16,18
X	DATABASE EMBL [Online] 18 November 1999 (1999-11-18) "Homo sapiens chromosome 3 clone RP11-195J23, working draft sequence, 32 unordered pieces" Database accession no. AC015547 XP002214755 the whole document ---	1
P,X	WO 01 04305 A (INCYTE GENOMICS INC ;YUE HENRY (US); TANG Y TOM (US)) 18 January 2001 (2001-01-18) SEQ ID NOS:1 and 3 ---	1,2,9, 10, 14-16, 18-20
X	DATABASE EMBL [Online] 8 June 2000 (2000-06-08) "EST369211 MAGE resequences, MAGD Homo sapiens cDNA, mRNA sequence." Database accession no. AW957141 XP002214756 the whole document ---	1,2
X	DATABASE EMBL [Online] 25 September 1999 (1999-09-25) "Homo sapiens mRNA; EST DKFZp761A1913_r1 (from clone DKFZp761A1913)" Database accession no. AL119221 XP002214757 the whole document ---	1,2
Y	GINSBERG STEPHEN D ET AL: "Expression profile of transcripts in Alzheimer's disease tangle-bearing CA1 neurons." ANNALS OF NEUROLOGY, vol. 48, no. 1, July 2000 (2000-07), pages 77-87, XP008008699 ISSN: 0364-5134 the whole document --- -/--	1,2,4-20

Form PCT/ISA210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT		Inter: at Application No PCT/US 01/21982
C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 96 14435 A (UNIV PENNSYLVANIA ;EBERWINE JAMES (US); DICHTER MARC (US); MIYASHI) 17 May 1996 (1996-05-17) the whole document	1,2,4-20
X	MIYAZAKI TAKASHI ET AL: "Molecular cloning of a novel thyroid hormone-responsive gene, ZAKI-4, in human skin fibroblasts" JOURNAL OF BIOLOGICAL CHEMISTRY, AMERICAN SOCIETY OF BIOLOGICAL CHEMISTS, BALTIMORE, MD, US, vol. 271, no. 24, 1996, pages 14567-14571, XP002152090 ISSN: 0021-9258 the whole document	1,2,7,8, 11,16
X	KINGSBURY TAMI J ET AL: "A conserved family of calcineurin regulators" GENES AND DEVELOPMENT, COLD SPRING HARBOR LABORATORY PRESS, NEW YORK, US, vol. 14, no. 13, 1 July 2000 (2000-07-01), pages 1595-1604, XP002212247 ISSN: 0890-9369 figure 2C	1,2,7,8
X	FUENTES JUAN J ET AL: "DSCR1, overexpressed in Down syndrome, is an inhibitor of calcineurin-mediated signaling pathways" HUMAN MOLECULAR GENETICS, OXFORD UNIVERSITY PRESS, SURREY, GB, vol. 9, no. 11, January 2000 (2000-01), pages 1681-1690, XP002212245 ISSN: 0964-6906 the whole document	1,2,7,8
A	GINSBERG S D ET AL: "Expression profile of individual Alzheimer's disease neurofibrillary tangle-bearing CA1 neurons using gene microarrays." SOCIETY FOR NEUROSCIENCE ABSTRACTS., vol. 25, no. 1-2, 1999, page 833 XP00115020 29th Annual Meeting of the Society for Neuroscience.;Miami Beach, Florida, USA; October 23-28, 1999 ISSN: 0190-5295 abstract 335.10	1,2,4-20
	--- -/--	

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

INTERNATIONAL SEARCH REPORT		Inten- Application No PCT/US 01/21982
C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SOMOGYI R ET AL: "Expression pattern of 9000 genes in Alzheimer's disease." SOCIETY FOR NEUROSCIENCE ABSTRACTS., vol. 25, no. 1-2, 1999, page 834 XP001108860 29th Annual Meeting of the Society for Neuroscience.;Miami Beach, Florida, USA; October 23-28, 1999 ISSN: 0190-5295 abstract 335.14 ---	1,2,4-20
A	FUHRMAN S ET AL: "Extracting meaningful patterns from functional genomics data in an animal Abeta toxicity model of Alzheimer disease." SOCIETY FOR NEUROSCIENCE ABSTRACTS., vol. 25, no. 1-2, 1999, page 834 XP001108859 29th Annual Meeting of the Society for Neuroscience.;Miami Beach, Florida, USA; October 23-28, 1999 ISSN: 0190-5295 abstract 335.12 ---	1,2,4-20
A	LORING JEANNE F ET AL: "Rational design of an animal model for Alzheimer's disease: Introduction of multiple human genomic transgenes to reproduce AD pathology in a rodent." NEUROBIOLOGY OF AGING, vol. 17, no. 2, 1996, pages 173-182. XP002215870 ISSN: 0197-4580 the whole document ---	10
A	STRIPPOLI PIERLUIGI ET AL: "A new gene family including DSCR1 (Down syndrome candidate region 1) and ZAKI-4: Characterization from yeast to human and identification of DSCR1-like 2, a novel human member (DSCR1L2)." GENOMICS, ACADEMIC PRESS, SAN DIEGO, US, vol. 64, no. 3, 15 March 2000 (2000-03-15), pages 252-263. XP002212248 ISSN: 0888-7543 the whole document ---	1,2,4-20
A	US 5 869 318 A (FUENTES JUAN JOSE ET AL) 9 February 1999 (1999-02-09) the whole document --- -/--	1,2,4-20

Form PCTISA210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT		Inter: # Application No PC1/US 01/21982
C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SUZUKI TAKASHI ET AL: "Molecular cloning of a novel apoptosis-related gene, human Nap1 (NCKAP1), and its possible relation to Alzheimer disease." GENOMICS, vol. 63, no. 2, 15 January 2000 (2000-01-15), pages 246-254, XP002215871 ISSN: 0888-7543 the whole document	1,2,4-20
A	YEGHIAZARYAN K ET AL: "Downregulation of the transcription factor scleraxis in brain of patients with Down Syndrome." JOURNAL OF NEURAL TRANSMISSION SUPPLEMENT, vol. 57, 1999, pages 305-314, XP008008700 ISSN: 0303-6995 the whole document	1,2,4-20
P,X	DATABASE BIOSIS [Online] BIOSCIENCES INFORMATION SERVICE, PHILADELPHIA, PA, US; 2000 LORING JEANNE F ET AL: "Molecular pathology: Gene expression microarrays used to analyze Alzheimer's disease brain." Database accession no. PREV200100494470 XP002215872 abstract & INTERNATIONAL GENOME SEQUENCING AND ANALYSIS CONFERENCE, vol. 12, 2000, pages 26-27, 12th International Genome Sequencing and Analysis Conference; Miami Beach, Florida, USA; September 12-15, 2000 -----	1,2,4-20

Form PCTISA210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT		International application No. PCT/US 01/21982
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.:
		1, 2, 4-20
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 01/21982

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

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

1. Claims: 1,2,4-20 all completely
Nucleic acid molecule and corresponding protein as defined by SEQ ID NOS:1 and 2, compositions, substrate, vectors, host cell, and methods based on said sequences.
2. Claim : 3 partially
cDNA molecule as defined by SEQ ID NO:3
3. Claim : 3 partially
cDNA molecule as defined by SEQ ID NO:4
4. Claim : 3 partially
cDNA molecule as defined by SEQ ID NO:5
5. Claim : 3 partially
cDNA molecule as defined by SEQ ID NO:6
6. Claim : 3 partially
cDNA molecule as defined by SEQ ID NO:7
7. Claim : 3 partially
cDNA molecule as defined by SEQ ID NO:8
8. Claim : 3 partially
cDNA molecule as defined by SEQ ID NO:9

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter: Application No
PCT/US 01/21982

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0194629	A	13-12-2001	AU 6455901 A 17-12-2001
			WO 0194629 A2 13-12-2001
			US 2002165180 A1 07-11-2002
			US 2002115057 A1 22-08-2002
			US 2002081301 A1 27-06-2002
			US 2002110821 A1 15-08-2002
			US 2002102531 A1 01-08-2002
			US 2002150877 A1 17-10-2002
			US 2002102532 A1 01-08-2002
			US 2002115085 A1 22-08-2002
			WO 0204491
AU 7330501 A 21-01-2002			
WO 0204491 A2 17-01-2002			
WO 0104305	A	18-01-2001	AU 6206500 A 30-01-2001
			EP 1196574 A2 17-04-2002
			WO 0104305 A2 18-01-2001
WO 9614435	A	17-05-1996	US 5723290 A 03-03-1998
			EP 0787206 A1 06-08-1997
			JP 10508490 T 25-08-1998
			WO 9614435 A1 17-05-1996
US 5869318	A	09-02-1999	ES 2154506 A1 01-04-2001

フロントページの続き

(51) Int.Cl. ⁷	F I	テーマコード(参考)
C 1 2 N 1/15	C 1 2 N 1/15	4 B 0 6 5
C 1 2 N 1/19	C 1 2 N 1/19	4 C 0 8 4
C 1 2 N 1/21	C 1 2 N 1/21	4 H 0 4 5
C 1 2 N 5/10	C 1 2 P 21/02	C
C 1 2 P 21/02	C 1 2 Q 1/68	A
C 1 2 Q 1/68	G 0 1 N 33/15	Z
G 0 1 N 33/15	G 0 1 N 33/50	Z
G 0 1 N 33/50	G 0 1 N 33/53	M
G 0 1 N 33/53	G 0 1 N 37/00	1 0 2
G 0 1 N 37/00	C 1 2 N 15/00	F
// A 6 1 K 45/00	C 1 2 N 5/00	A
A 6 1 P 25/28	A 6 1 K 45/00	
	A 6 1 P 25/28	

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

(72) 発明者 ティングリー、デボラ・ダブリュ

アメリカ合衆国カリフォルニア州 9 4 1 3 1 ・ サンフランシスコ ・ ノルビューウェイ 1 1 9

(72) 発明者 エドワーズ、カーラ・エム

アメリカ合衆国カリフォルニア州 9 4 0 1 9 ・ ハーフムーンベイ ・ ベンチュラストリート 4 1

(72) 発明者 ストリーター、デイビッド・ジー

アメリカ合衆国カリフォルニア州 9 5 0 0 6 ・ ボールドーククリーク ・ パインクレストドライブ 1 2 1 1

F ターム(参考) 2G045 AA25 AA29 AA40 BB03 BB20 CB01 CB09 CB16 CB17 DA12
 DA13 DA14 DA36 DA37 DA77 FB02 FB03
 4B024 AA01 AA11 BA80 CA01 CA04 DA02 DA05 DA11 EA02 EA03
 EA04 FA02 GA11 HA03 HA12
 4B029 AA07 BB20 CC03 FA12
 4B063 QA01 QA18 QA19 QQ05 QQ13 QQ42 QQ52 QQ79 QR08 QR42
 QR55 QR62 QR82 QS25 QS36 QX02
 4B064 AG01 CA01 CA19 CC24 DA01 DA13
 4B065 AA01X AA57X AA90X AA93Y AB01 BA01 CA24 CA25 CA44 CA46
 4C084 AA17 NA14 ZA012 ZA152 ZA162
 4H045 AA10 AA11 AA20 AA30 BA10 BA41 DA00 DA76 EA20 EA50
 FA72 FA74

专利名称(译)	唐氏综合症重要区域1 - 像1蛋白质		
公开(公告)号	JP2004516815A	公开(公告)日	2004-06-10
申请号	JP2002509376	申请日	2001-07-11
[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	ローリング・ジーン・エフ ティングリー・デボラ・ダブリュ エドワーズ・カーラ・エム ストリーター・デイビッド・ジー		
发明人	ローリング、ジーン・エフ ティングリー、デボラ・ダブリュ エドワーズ、カーラ・エム ストリーター、デイビッド・ジー		
IPC分类号	A01K67/027 A61K38/00 A61K45/00 A61K48/00 A61P25/28 C07K14/47 C07K16/18 C12M1/00 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N15/09 C12N15/12 C12P21/02 C12Q1/68 G01N33/15 G01N33/50 G01N33/53 G01N37/00		
CPC分类号	A61K38/00 A61K48/00 A61P25/28 C07K14/47 C12Q1/6883 C12Q2600/158		
FI分类号	C12N15/00.ZNA.A A01K67/027 C07K14/47 C07K16/18 C12M1/00.A C12N1/15 C12N1/19 C12N1/21 C12P21/02.C C12Q1/68.A G01N33/15.Z G01N33/50.Z G01N33/53.M G01N37/00.102 C12N15/00.F C12N5/00.A A61K45/00 A61P25/28		
F-TERM分类号	2G045/AA25 2G045/AA29 2G045/AA40 2G045/BB03 2G045/BB20 2G045/CB01 2G045/CB09 2G045/CB16 2G045/CB17 2G045/DA12 2G045/DA13 2G045/DA14 2G045/DA36 2G045/DA37 2G045/DA77 2G045/FB02 2G045/FB03 4B024/AA01 4B024/AA11 4B024/BA80 4B024/CA01 4B024/CA04 4B024/DA02 4B024/DA05 4B024/DA11 4B024/EA02 4B024/EA03 4B024/EA04 4B024/FA02 4B024/GA11 4B024/HA03 4B024/HA12 4B029/AA07 4B029/BB20 4B029/CC03 4B029/FA12 4B063/QA01 4B063/QA18 4B063/QA19 4B063/QQ05 4B063/QQ13 4B063/QQ42 4B063/QQ52 4B063/QQ79 4B063/QR08 4B063/QR42 4B063/QR55 4B063/QR62 4B063/QR82 4B063/QS25 4B063/QS36 4B063/QX02 4B064/AG01 4B064/CA01 4B064/CA19 4B064/CC24 4B064/DA01 4B064/DA13 4B065/AA01X 4B065/AA57X 4B065/AA90X 4B065/AA93Y 4B065/AB01 4B065/BA01 4B065/CA24 4B065/CA25 4B065/CA44 4B065/CA46 4C084/AA17 4C084/NA14 4C084/ZA012 4C084/ZA152 4C084/ZA162 4H045/AA10 4H045/AA11 4H045/AA20 4H045/AA30 4H045/BA10 4H045/BA41 4H045/DA00 4H045/DA76 4H045/EA20 4H045/EA50 4H045/FA72 4H045/FA74		
优先权	09/614474 2000-07-11 US		
外部链接	Espacenet		

摘要(译)

本发明提供核酸分子及其片段。本发明还提供了使用该核酸分子表征，诊断，评估和治疗与基因表达有关的症状，疾病和病症的方法，以及制备模型系统的方法。本发明进一步提供了表达载体和宿主细胞，用于产生由核酸分子编码的蛋白质。

